UNIVERSITY OF MIAMI
NOV 3 0 1965
LIBRARY

industrial mathematics

1954

UNIVERSITY OF MIAMI
NOV 3 0 1965
LIBRARY

industrial mathematics

1954

EDITORIAL STAFF

F. L. MONCHER, EDITOR-IN-CHIEF, VICKERS INCORPORATED
A. S. BEAM, VINCO CORPORATION
R. V. CHURCHILL, UNIVERSITY OF MICHIGAN
K. W. FOLLEY, WAYNE UNIVERSITY
A. GEMANT, DETROIT EDISON COMPANY
D. HUGHSON, PACKARD MOTOR CAR COMPANY
A. W. JACOBSON, WAYNE UNIVERSITY
E. D. McCARTHY, UNIVERSITY OF DETROIT
C. A. PIPER, BENDIX RESEARCH LABORATORY
A. E. ROACH, GENERAL MOTORS CORPORATION
R. A. ROGGENBUCK, FORD MOTOR COMPANY
W. F. VOGEL, WAYNE UNIVERSITY

VOLUME 5

Copyright 1955 by

INDUSTRIAL MATHEMATICS SOCIETY
100 FARNSWORTH AVENUE, DETROIT 2, MICHIGAN

The Industrial Mathematics Society wishes to express its sincere appreciation to the International Business Machines, Vinco Corporation, and Studebaker - Packard Corporation for providing financial assistance in the publication of Volume 5 of this Journal.

NUCLEAR ENERGY IN INDUSTRY

HENRY J. GOMBERG
Associate Professor of Electrical Engineering
and Assistant Director, Michigan Memorial-Phoenix Project

A nuclear energy industry, as part of the productive economy of our country, is practically nonexistent. The industry as does exist is almost entirely the property of the United States Government and is directed toward particular purposes: armament and defense. Likewise, the utilization of nuclear energy in industrial processes is extremely limited. However, the potential uses for nuclear energy in industry and the possibilities of a nuclear energy industry are so vast that this, as a general field of research, is occupying a very large part of our time and attention.

In analyzing the field as a whole, we have adopted the custom of breaking it up into three technical areas: nuclear power, radiation, and tracers. Nuclear power is the basis for a nuclear energy industry as such. The others present great possibilities for utilization of nuclear energy in industry.

Fundamentally, we are dealing with a new source of energy — so new that here we can really talk about a "first" for man. If we go back and look at all the energy sources that were available to us throughout the history of civilization, we find that they inevitably trace their existence back to the sun. Coal, oil, gas, water power, wind, or any of the "Rube Goldberg" schemes that have occasionally been proposed or tried, eventially trace their ultimate energy availability to the sun. In nuclear energy, for the first time, we have a process in the same class as that which takes place within the sun itself. It is not identical with it, but we have reason to believe that the energy in the sun is nuclear energy. For the first time, we have an energy source which is independent of the sun.

But how important is it? Not very many years ago some very prominent people in the field of physics and economics were warning us not to get too excited about the development of atomic energy because the total resources available in the nuclear fuels are too small to be particularly significant. We know approximately how much coal is available in the earth; and if we account for all that which can be mined by processes costing up to about four times the present figure, we find that there is enough coal to produce 722×10^{17} British thermal units of heat.

When we examine the older ideas of the availability of atomic energy in the form of uranium 235 only, we find there is just about 1/7 the above amount in all the known uranium deposits in the world. While it may be interesting, this is not really exciting, and it does not impress us too much as something which is going to change our future economy. But if we take a closer look, we find that the mathematics of nuclear fission indicate, by a process known as breeding, the factor of 1/7 the known resources of fuel can jump to about 20 times all the known resources of fuel. Uranium 235, the basis of the original calculation, is a fairly rare material. It occurs once in every 143 atoms of natural uranium. The rest is uranium 238. But this uranium 238, under the proper conditions, can be converted to a fissionable material in the so-called breeder reactor. This means that by the breeding process,* all the 142 atoms which we literally throw away at such places as the gaseous diffusion plants at Oak Ridge, could eventually be put to use. It is conceivable that we will convert this now

*The Experimental Breeder Reactor at Arco, Idaho, has operated on this principle since 1951.

wasted material into a new form of fuel. There have been many stories told about the possibilities of these breeder reactors. The use of analogies is always dangerous, but sometimes they help to clarify the picture, so we will try one. The breeder reactor is like a furnace which would be filled at the beginning of the winter with stones and a little coal. The house would be heated all winter, and then a furnace full of coal dug out in the spring. In other words, when starting out with a material that is of no particular use for heating, heat extraction will take place, the inert material will be converted to a new fuel.

If we move in this new direction, then we immediately have to ask questions as to what is going to happen in industry and to our general standard of living when very large new sources of energy become available. I believe Professor Leslie White, on this campus (University of Michigan), has tried to estimate levels for the measurement of civilizations as to what basis would be used to determine the advancement of a particular group of people as a whole. One of the yardsticks that has been proposed is a measure of how much horsepower and how much machinery are available per capita to do the work required to keep the civilization going. In other words, how much is available to everybody — not just the small group on top, but to every one individually. It turns out that for the United States, the figure is about 125 horsepower per capita.

But this 125 horsepower per capita is dependent entirely on the fuel resources that are available to keep them going. I have some more figures that I would like to cite because they are rather startling when read. These are figures developed by the Bureau of Mines, which estimate that with all the possible fuels available in the earth's surface, with conversion and all kinds of economy measures being taken, there is a fuel supply for 650 years at the present rate of consumption. (This excludes the nuclear fuels). If we allow for the normal population expansion that has been going on in the United States and also the rate of increase in power consumption that has been going on, the supply period comes down to about 150 years. This is altogether too small a margin to live with comfortably. The same margin existed ten years ago just as it exists today, but now it is believed feasible to utilize nuclear fuels in a breeder or convertor process, and in this way, the 150 year reserve has been increased by a factor of 20. We are now talking about some 3000 years and this is a long enough period so that we need not worry about it.

Another very important aspect is the form of the stored energy. Is this just another lump of coal, or are we dealing with a fundamentally new kind of fuel? As I mentioned a while ago, this is the only process known to man which he can control at the present time and which is competitive with the reactions that take place in the sun. A unique feature of nuclear fuels is the small package which contains the energy. It has been said with complete justification that if all the heat in a kilogram of uranium 235 (about 2 pounds) were released, the heat resulting would be equivalent to 20,000 tons of coal — or pictorially, 400 coal cars, each carrying about 50 tons of coal. This kilogram of uranium would probably be smaller than a tennis ball. The use of nuclear energy results in fuel packages of this order of magnitude, whereas the nearest equivalent would fill a freight train four miles long. So, we have not only a new source of fuel that will keep things going, but we have a new fuel, which in its future development, in its engineering potentialities, promises startling developments. It is a completely new form of energy, and it is a new phenomen which we must learn to deal with and to work with.

Through the Phoenix Project (University of Michigan) and the industrial participation program of the Phoenix Project, I think a very large step forward has been taken in the direction of learning to live with this new material.

Through a grant of the Ford Motor Company, we are now in a position to build our own nuclear reactor. This will be the first moderate size nuclear reactor on any university campus.

We have been able to develop, again through the help of the laboratories and the facilities of the project, educational opportunities in this new area, and we are now offering a graduate

degree in nuclear engineering. This would not seem to be a very important development except that one of the most critical situations in the country at the present time is the lack of technical manpower. The lack of technical manpower in the nuclear field is more acute than in any other – and for a basic reason. In other fields we know what we can do and how many men are needed to do it; but if just one of the developments resulting from the research in nuclear energy should go into industry, the need for trained personnel is going to skyrocket.

Going one step further now, what else is going on in the nuclear field? We have entered it from an educational point of view and we have entered it from a research point of view. How about the actual business of trying to get this nuclear energy out to where we can make use of it? As you may know, the Detroit Edison Company, by means of the Atomic Power Development Associates, is practically leading the field in a major effort to bring electricity, generated from nuclear fuels into the practical application stage in as short a time as possible. In other words, they are trying to supplement or supplant coal as a fuel and use the heat that is available in the nuclear reaction. At the same time there is thought of making more plutonium available. But one of the most interesting things about the Detroit development is that it is being designed with the ultimate aim of producing plutonium which will not necessarily have the tremendous markup it has now as bomb material. It could go back into a fuel stockpile, and still the process would be competitive with coal utilization. This would not mean generation of power on a government subsidy because we are trying to manufacture weapons; but it would mean manufacture of power on a competitive, economic, and everyday basis.

I can tell you that the University is a participating group. Our participation is expanding as our laboratories expand; and as our supply of technical personnel increases, we hope to play a larger and larger part in that development.

So much for the field of nuclear power itself. Let us now examine radiation and tracer work. In radiation we again have a new kind of energy. The major sources of radiation in the form of high energy electromagnetic waves and particles are the nuclear reactors themselves, and also the fission products or the pieces that are left over after uranium fission has taken place. The fission fragments actually carry in them a fairly large amount of energy. There are about 200,000,000 electron volts available in an atom that is about to split. Of this, about 160,000,000 go into the actual fission process. This would be the major part of the heating and energy liberated, let us say, in a nuclear power plant. The other 40,000,000 from the original particle are contained in the leftover pieces and are now available for other types of work. Some of these other types of work promise to be as interesting and in some cases as important as the energy from fission. About half of the energy left is in a form equivalent to X-ray machines. The radioactive cesium, which will be used in the form of an irradiator in the Project's Alice Lloyd Radiation Therapy Laboratory, is a fission product.

These forms of energy were once so expensive that they could only be considered in fields where cost is no object – as in saving peoples' lives. Now the cost is beginning to come down so that we can consider these sources of energy for industrial processes. When we consider them in this way, the question comes up as to what they are good for. It is amazing how little we know about what can be done with them. This is one of the major areas of activity in the Phoenix Project – to find out what can be done with this high energy radiation on the assumption that it is available in large quantities.

One of the things being investigated is the promotion of chemical reactions. It is known that certain reactions which normally go very slowly can be accelerated in the presence of high energy radiation.

I would like to discuss only things we have done rather than to speculate, so I will discuss one project, and that is the work being done on the production of lindane. Lindane is an important insecticide and is manufactured by adding chlorine to benzene (the additive chlorination of benzene). This takes place now industrially by subjecting the materials to ultraviolet

radiation. Ultraviolet radiation is very easily absorbed since it will not penetrate any great distance through matter. As a result, the reaction has to be carried on with the materials spread out in very thin sheets. Since the reaction takes place rather rapidly, this is not a tremendous disadvantage; but nevertheless, it is one of the limitations of the process. However, through the use of gamma radiation, which can penetrate right through the steel walls of the vessel, it is possible to put the radiation source outside and have the reaction take place in bulk inside the vessel. In this manner, lindane has been manufactured successfully in our laboratories.

More interesting is the possibility that a new material has been made. There is nothing reported in the literature on the addition of chlorine to toluene, a chemical compound related to benzene. So far as can be told in all the tests made, there has been successful chlorination of toluene under gamma irradiation. What is it good for? Nobody knows or is even sure of it. But I would like to point out that every time a new product, a new material, and a new method of making things have been developed, literally, the roof blows off somewhere in science and technology. The present rapid expansion of the plastics industry and the synthetic rubber industry are typical examples of what happens when a little bit more is learned about how to do things. We fully anticipate similar development in industrial processes through the use of radiation.

Another area in which we are very much interested, a field in which Dr. Gould of the Wayne County General Hospital and I have been working now for about a year, is the field of disease control through radiation. The specific disease under study is trichinosis, concerning which Dr. Gould is recognized as an international authority. In this particular problem we are studying a parasitic disease cycle in which the parasite exists in pork. If the infected material is eaten without being properly cooked, the disease is contacted. If some of the infected material is passed along in the form of garbage to a pig, he also contacts the disease. The degree of infection depends on how many of the trichina were ingested, and the severity of the symptoms depends somewhat on the individual. One of the most important aspects of this particular disease is that it works on the basis of multiplication. For every three trichina (two females to every male is about the population ratio) that get into the body, 3000 more will be produced. It is these 3000 that we really worry about, for they are the major infective agents in this disease.

It turns out that, through the use of radiation, one of three things can be done. We can irradiate the food very heavily and kill all the trichina larvae in the meat. It is perfectly safe, but it cannot be eaten because the pork tastes terrible. On the other hand, we can cut this dose by a factor of about 50 and it turns out that the trichina larvae that emerge in the stomach and then go to the intestinal tract will never develop. They will not mature to adults and therefore cannot reproduce. Finally, we can do something very much sinister by cutting the dose about one third again. The trichina will enter the body with the food, will move down to the intestines, will grow up, and the males and females will get together, but they are now wasting time because the females are all sterile. The reproductive phase of the disease has been destroyed.

Leaving the uses of radiation, I would like to discuss another field — namely, the use of tracers. Here the most important aspect of the work is research, and certainly in the University we are extremely research conscious. Tracers provide us with the opportunity of getting new information and, in that sense, may be more important than any of the other things herein discussed. The radioactive cesium that was mentioned, the radioactive iodine, and radioactive phosphorus are atoms having a rather peculiar life history. For instance, the iodine atoms go along for a certain period of time behaving as any iodine atoms should. But, at an average rate of about eight days for a group, half of them will convert themselves to xenon gas. Now fortunately, in the use of radioactive iodine, the development of this gas does not present any sort of hazard because the amount of gas involved is too small to be serious. But the important thing is that there is a change in chemical form and this change is accompanied by the emission of radiation.

A new diagnostic technique for locating brain tumors is based on the fact that when iodine is incorporated in a brain tumor, the radiation leaving it can be detected outside the body through the use of proper instruments. In other words, it is now possible to trace the atoms in their paths by a chemical reaction or in a physical system or in any situation considered. It is possible to trace them wherever they go.

So, we can run the gamut through power, radiation, and tracers. The Phoenix Project is in all three fields which are very important to industry and society.

DERIVATION OF THE FUNDAMENTAL EQUATION OF RESIDUAL STRESS ANALYSIS BY DISSECTION

LEONARD G. JOHNSON
Mathematician; Engineering Mechanics Department
Research Laboratories Division, General Motors Corporation
Detroit, Michigan

By starting with such assumptions as are usually made in uniaxial residual stress analysis by dissection, the author arrives at a fundamental equation which is completely general, and which may be used in determining residual stresses in any uniaxial problem in any curvilinear coordinate system. The general equation is verified as correct for the special case of straight beams, since it reduces to the standard formula for the longitudinal residual stress in such beams as obtained by observing curvature changes.

INTRODUCTION

Every exact science consists essentially of two parts:

(a) The so-called basic assumptions, or postulates, of the science.

(b) The necessary consequences, or theorems, which follow from the postulates according to mathematical logic.

In this discussion of the dissection method of determining longitudinal residual stresses in bars, we shall first of all give a list of postulates on which the entire scheme of analysis is built, and then proceed to derive the fundamental equation by manipulating the postulated equations in accordance with the rules of elementary mathematics. It will then become evident that the fundamental equation of the dissection process, although profound in appearance, is not so profound after all.

POSTULATES OF THE DISSECTION PROCESS

POSTULATE I: The change in stress produced at any location by a cut in directly proportional to the total force which existed in the layer removed. That is, if a surface stress σ_0 existed in the layer dZ which is removed, then the stress change at any location x units below the original top is

$$d\sigma_x = g_x \sigma_0 \, dZ$$

POSTULATE II: The proportionality factor, g_x, of postulate I is fixed for a fixed geometry, being independent of the longitudinal stress distribution existing in the specimen, as long as we are dealing with the same location at the same stage of cutting.

POSTULATE III: The stress at location x when nothing has been removed is equal to the stress at location x when the amount x has been removed minus the cumulative stress change at location x due to the removal of successive layers amounting to a total of x.

EXPLANATION OF THE ADOPTED SYMBOLISM

In order to make clear the meanings of the symbols to be used in our general analysis of the dissection process, let us refer to Figure 1.

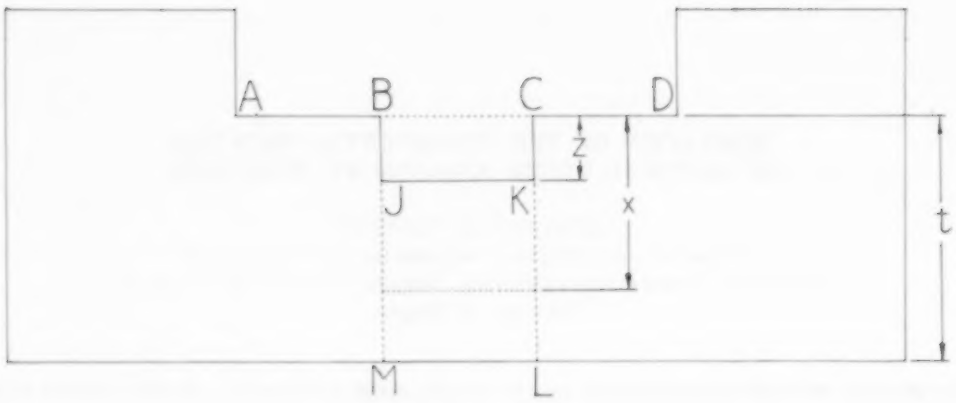


Figure 1

A specimen whose original top surface is ABCD is to be dissected in the region between BM and CL. The letter t denotes the initial thickness in said region. The letter x is a locating dimension from the original top to whatever level we are interested in. In order to avoid confusion, we shall put the subscript L on the letter x and made it read x_L, the letter L being the first letter of the word locating. Thus, x_L will always denote a locating dimension of magnitude x. The letter Z shall denote the amount of stock which has been removed. Hence, at the beginning of the process Z = 0. In order to emphasize that Z is a removal dimension, we shall put the subscript r on the letter Z and make it read Z_r. Thus, Z_r denotes a removal dimension of magnitude Z, the letter r being the first letter of the word removal.

The general symbol for the longitudinal stress at any level at any stage shall be $\sigma_x(Z)$, the subscript x being the locating dimension to the level in which we are interested, and the argument Z denoting the amount of stock already removed. Thus, the longitudinal stress at the level x when an amount Z has been removed shall be denoted by the symbol $\sigma_{xL}(Z_r)$.

DERIVATION OF THE FUNDAMENTAL EQUATION

In terms of the symbolism just explained, we may write postulate III as follows:

$$\int_{x_r}^{0} d\sigma_{x_L}(Z_r) - \sigma_{x_L}(x) = \sigma_{x_L}(0) \tag{1}$$

According to postulate I, we have

$$d\sigma_{x_L}(Z_r) = g_{x_L}(Z_r) \cdot \sigma_{Z_L}(Z_r) \cdot dZ_r \tag{2}$$

where $g_{x_L}(Z)$ is the change of stress at location x per unit of additional force removed at the top when an amount of stock Z has already been removed. This quantity $g_{x_L}(Z_r)$ shall be known as the influence factor at x when Z has been removed.

(2) may be put into the form

$$\sigma_{Z_L}(Z_r) = \frac{\sigma'_{x_L}(Z_r)}{\varepsilon_{x_L}(Z_r)} \tag{3}$$

where $\sigma'_{x_L}(Z_r)$ is the derivative of $\sigma_{x_L}(Z_r)$ with respect to Z_r.

(3) holds for all values of x_L and Z_r. Hence, we may put $x_L = t$ and $Z_r = x_r$. Then

$$\sigma_{x_L}(x_r) = \frac{\sigma'_t(x_r)}{\varepsilon_t(x_r)} \tag{4}$$

We may therefore rewrite (1) in the form

$$\sigma_{x_L}(0) = \frac{\sigma'_t(x_r)}{\varepsilon_t(x_r)} - \int_0^{x_r} \sigma'_{x_L}(Z_r)\, dZ_r \tag{5}$$

According to (3), we have

$$\sigma'_{x_L}(Z_r) = \varepsilon_{x_L}(Z_r) \cdot \sigma_{Z_L}(Z_r) \tag{6}$$

We may put $x_L = t$ in (3). Then we obtain

$$\sigma_{Z_L}(Z_r) = \frac{\sigma'_t(Z_r)}{\varepsilon_t(Z_r)} \tag{7}$$

Substituting the right member of (7) for $\sigma_{Z_L}(Z_r)$ in (6), we obtain

$$\sigma'_{x_L}(Z_r) = \frac{\varepsilon_{x_L}(Z_r) \cdot \sigma'_t(Z_r)}{\varepsilon_t(Z_r)} \tag{8}$$

Therefore, (5) becomes

$$\sigma_{x_L}(0) = \frac{\sigma'_t(x_r)}{\varepsilon_t(x_r)} - \int_0^{x_r} \frac{\varepsilon_{x_L}(Z_r) \cdot \sigma'_t(Z_r)}{\varepsilon_t(Z_r)}\, dZ_r \tag{9}$$

In terms of longitudinal strains, this may be written as

$$\frac{\sigma_{x_L}(0)}{E} = \frac{\epsilon'_t(x_r)}{\varepsilon_t(x_r)} - \int_0^{x_r} \frac{\varepsilon_{x_L}(Z_r) \cdot \epsilon'_t(Z_r)}{\varepsilon_t(Z_r)}\, dZ_r \tag{10}$$

where E is Young's modulus.

The strain derivative $\epsilon'_t(x_r)$, at the bottom, can be determined experimentally on the specimen by placing a strain gauge at the bottom and observing the change in strain at each step. Since all that is required is the derivative, the initial value of the strain is immaterial. The influence factors $g_{x_L}(Z_r)$ and $g_t(Z_r)$ may be determined either theoretically, or experimentally, from a geometrically similar model, as will be explained in the next section. We have thus derived the fundamental equation (10) for residual stress analysis by dissection. Dropping the identification subscripts L and r, (10) becomes

$$\frac{\sigma_{x(0)}}{E} = \frac{\epsilon'_t(x)}{g_t(x)} - \int_0^x \frac{g_x(Z) \cdot \epsilon'_t(Z)}{g_t(Z)} dZ \qquad (11)$$

DETERMINING THE INFLUENCE FACTOR $g_x(Z)$

In the most general type of problem, the geometry of the specimen will make the theoretical determination of the influence factor a prohibitive task. Hence, in such a case it will be preferable to determine the influence factor experimentally by applying some type of loading to a large scale model which is geometrically similar to the actual specimen. Successive layers may be removed from the model, and the stress changes at any location x can be determined by observing a strain gauge reading at x before and after each removal while the same external loading is maintained on the model. This is all based on postulate II, which states that the influence factors are functions of geometry only.

AN EXAMPLE OF THE THEORETICAL DERIVATION OF THE INFLUENCE FACTOR

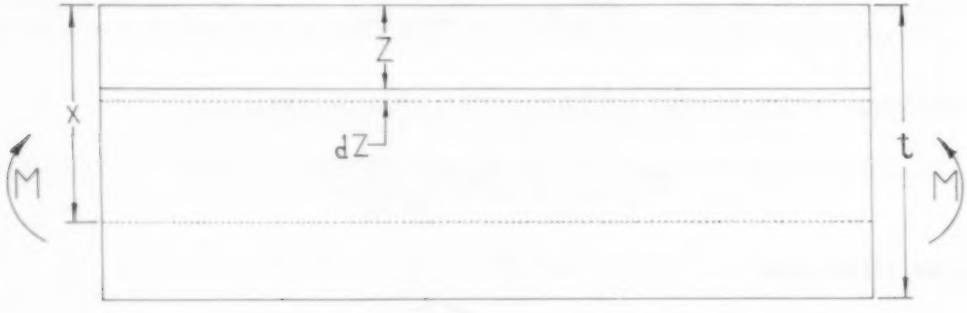


Figure 2

Consider a flat rectangular bar (figure 2) of original thickness t. Suppose an amount of stock Z has already been removed from the top. The stress at location x below the original top due to a pure bending moment M is given by the classical flexure formula [1][1]:

$$\sigma_x = \frac{\text{(Bending moment)}\cdot\text{(Distance from x to Neutral Axis)}}{\text{Moment of Inertia of Section of Thickness (t - Z)}}$$

[1] Numbers in brackets refer to the bibliography at the end of the paper.

In terms of the notation indicated in figure 2, this longitudinal bending stress is

$$\sigma_x = \frac{M\left[1 - \frac{2(x-Z)}{t-Z}\right]}{\frac{(t-Z)^2}{6}}$$

The dimension of the bar perpendicular to the page is taken as unity.

Thus,

$$\sigma_x = 6M\left[\frac{1}{(t-Z)^2} - \frac{2(x-Z)}{(t-Z)^3}\right] \tag{12}$$

The influence factor is defined mathematically in postulate I as being

$$g_x = \frac{d\sigma_x}{dZ} \Big/ \sigma_o = \frac{\sigma'_x}{\sigma_o}$$

where σ_0 represents the longitudinal stress at the top surface where x = Z. Hence,

$$\sigma_o = \frac{6M}{(t-Z)^2} \tag{13}$$

Differentiating (12) with respect to Z we obtain

$$\frac{d\sigma_x}{dZ} = 6M\left[\frac{4}{(t-Z)^3} - \frac{6(x-Z)}{(t-Z)^4}\right] \tag{14}$$

Hence, the desired influence factor is

$$g_x(Z) = \frac{d\sigma_x}{dZ} \Big/ \sigma_o = \frac{4}{t-Z} - \frac{6(x-Z)}{(t-Z)^2} \tag{15}$$

$$= \left(\frac{1}{t-Z}\right)\left[4 - 6\left(\frac{x-Z}{t-Z}\right)\right]$$

(11) can now be used to determine longitudinal residual stresses in flat rectangular bars by substituting (15) for $g_x(Z)$, as shown in the next section.

THE UNIAXIAL RESIDUAL STRESS FORMULA FOR STRAIGHT RECTANGULAR BARS

We start with the fundamental equation (11):

$$\frac{\sigma_x(0)}{E} = \frac{\epsilon'_t(x)}{g_t(x)} - \int_0^x \frac{g_x(Z)\cdot\epsilon'_t(Z)}{g_t(Z)}\,dZ$$

The influence factor $g_x(Z)$ is given by (15). The function $g_t(Z)$ is what (15) becomes at $x = t$, i.e.,

$$g_t(Z) = \left(\frac{1}{t - Z}\right)(4 - 6) = -\frac{2}{t - Z} \tag{16}$$

Substituting (15) and (16) into the fundamental equation, we obtain

$$\frac{\sigma_x(0)}{E} = -\left(\frac{t - x}{2}\right)\epsilon_t'(x) + \int_0^x \left[2 - 3\left(\frac{x - Z}{t - Z}\right)\right] \epsilon_t'(Z)\, dZ \tag{17}$$

Integrating by parts, taking $u = 2 - 3(\frac{x - Z}{t - Z})$ and $dv = \epsilon_t'(Z)\, dZ$, we find that (taking $\epsilon_t(0) = 0$),

$$\int_0^x \left[2 - 3\left(\frac{x - Z}{t - Z}\right)\right] \epsilon_t'(Z)\, dZ = 2\epsilon_t(x) - 3(t - x)\int_0^x \frac{\epsilon_t(Z)}{(t - Z)^2}\, dZ$$

Therefore,

$$\frac{\sigma_x(0)}{E} = -\left(\frac{t - x}{2}\right)\epsilon_t'(x) + 2\epsilon_t(x) - 3(t - x)\int_0^x \frac{\epsilon_t(Z)\,dZ}{(t - Z)^2} \tag{18}$$

To express (17) in terms of curvature changes, we use the basic relation that

$$\text{CURVATURE CHANGE} = \frac{\text{STRAIN CHANGE AT x}}{\text{DISTANCE FROM x TO NEUTRAL AXIS}}. \tag{19}$$

In (15) we see that the neutral axis is located at 2/3 of the remaining thickness below the top at any step. Hence, (18) may be written symbolically as

$$d\,c(Z) = \frac{d\epsilon_t(Z)}{(t - Z)/3}, \text{ or } \epsilon_t'(Z) = \left(\frac{t - Z}{3}\right) c'(Z) \tag{20}$$

where primes denote differentiation with respect to Z.

Substituting (20) for $\epsilon_t'(Z)$ in (17):

$$\frac{\sigma_x(0)}{E} = -\frac{(t - x)^2}{6} c'(x) + \frac{1}{3}\int_0^x (2t + Z - 3x)\, c'(Z)\, dZ \tag{21}$$

Integrating by parts, taking $u = 2t + Z - 3x$, and $dv = C'(Z)\, dZ$, we find that (taking $C(0) = 0$),

$$\int_0^x (2t + Z - 3x)\, c'(Z)dZ = 2(t - x)\, c(x) - \int_0^x c(Z)\, dZ$$

Therefore,

$$\frac{\sigma_x(0)}{E} = -\frac{(t-x)^2}{6}C'(x) + \frac{2}{3}(t-x)\,C(x) - \frac{1}{3}\int_0^x C(Z)\,dZ \qquad (22)$$

which is the standard uniaxial residual stress formula for straight beams in terms of observed curvature changes, C(x), as obtained by various authors [2], [3].

CONCLUSION

By starting with such assumptions as are usually made in uniaxial residual stress analysis by dissection, we have arrived at a fundamental equation which is completely general, and which may be applied to any uniaxial problem in any curvilinear coordinate system, once the so-called "influence factor" has been determined. The fundamental equation is found to be verified in the case of straight beams, since it reduces to the standard equation used in straight beam longitudinal residual stress analysis. The author has also computed influence factors for curved beams, and it is hoped that these results will be published in the near future.

BIBLIOGRAPHY

1. Seely, F. B.: Resistance of Materials; John Wiley & Sons, 1935; p. 82, eqn (35).
2. Stäblein, F.: "Spannungsmessungen an Einseitig Abgeloschen Knuppeln," Kruppsche Monatshefte 12, p. 93-98 (1931).
3. Letner, H. R.: "Application of Optical Interference to the Study of Residual Surface Stresses," Proceedings, Society for Experimental Stress Analysis, Vol. 10, No. 2, p. 26, Eqn (2), (1952).

POWER FOR GUIDED MISSILES AND SPACE SHIPS

MAGNUS VON BRAUN
Redstone Arsenal
Huntsville, Alabama

1. Introduction

Guided Missiles and Space Travel! These two words have a fascinating sound to Junior watching "Space Cadet" on television or reading about the adventures of Buck Rogers. To the average layman, they are intriguing subjects, often with a touch of something like black magic. But to the engineer associated with this work, the two words encompass a variety of scientific and engineering work in many branches of natural science, with a host of interesting and sometimes perplexing, but solvable problems; for them, work on guided missiles and space travel is very much down to earth, if I may use that paradox.

Development work on guided missiles and, eventually, space ships is and will be done in mechanical and structural design, electronics, physics, chemistry and medicine for manned ships and many other fields, but one of the most important branches of this development work is dedicated to the propulsion of these craft. And the speed requirements in guided missiles for military reasons, in space ships because the earth's gravitational force must be overcome, these speed requirements are so far beyond those of conventional aircraft that radically new approaches had to be taken.

The solution was found in jet propulsion whose development proceeded at a rapid rate under the stimulus of the Second World War. The public began to hear about V-bombs, rockets, jet-propelled aircraft and guided missiles; and, as a result, personnel in the fields of aviation and ordnance, as well as laymen have shown a keen interest in learning more about this new method of propulsion. In the following discussion, I have not attempted to present to you a comprehensive coverage of all aspects and types of jet propulsion, which would be far beyond the scope of such a paper, but rather to point out some of the problems involved in propelling guided missiles and, eventually, space ships at the high flight velocities required for these craft.

2. Fundamentals

The basis of jet propulsion is Newton's Third Law which states that for every action there is an equal and opposite reaction. Thus, if we want to move a body in one direction, we must have a force acting in the opposite direction. I only need to remind you of the recoil of a gun to illustrate what I am talking about. The powder charge exploding in the chamber of the gun exerts a pressure in all directions; the pressure acts on the bullet as much as it does all surfaces of the chamber, only the bullet can move and so it is pushed out. But the pressure acting in the opposite direction hits a blank wall and is transmitted to your shoulder. The larger the powder charge, the farther the bullet will go, the more recoil you will feel. Or, take the airscrew. By spinning a propeller, we exert a force on the air toward the rear, and this action causes the propeller, and with it the airplane, to move forward.

In jet propulsion, we do essentially the same thing: we push masses of gas rearward; and by doing so, we propel the body from which these gases emerge, forward. For a simplified

discussion of the problem, let us assume some body of a certain mass moving along at a certain velocity. We could keep this body moving if we ejected from it gases of the same mass traveling at the same velocity as that of the body we are trying to propel. But, since the impulse that we must transmit to our moving body is the product of mass and velocity of the gases we eject, we can play certain tricks. We are obviously limited with the mass of gas we can store or set free on board our body, but we can make up for this lack of mass by ejecting our gases at a higher velocity. And this is one of the primary goals of all development work in the field of jet propulsion: a high gas exhaust velocity.

We know from fundamental thermodynamics that in order to obtain high gas flow velocities, we must expand from high pressure to low pressure, that is, convert pressure or potential energy into velocity or kinetic energy; in other words, we need a high pressure ratio between the container from which we eject our gases, and the ambient air into which we eject, where the ambient air pressure depends, of course, upon the altitude at which our body travels. Our problem then boils down to building up and maintaining a high gas pressure in our container, high compared to the ambient atmospheric pressure. This pressure can, of course, best be obtained by increasing the volume of a mass of gas in question, and that obviously means raising its temperature by some kind of combustion. In simple terms, then, we can say that the hotter we can burn, the more we can increase the gas volume and with it the pressure in our container, thus producing a higher pressure ratio and a higher exhaust velocity of the gases as they are ejected from the container.

So, essentially, what we need for jet propulsion is some kind of a combustion chamber with a continuous supply of fuel and oxygen, and an exhaust opening pointing in the direction opposite to the desired flight direction. This, at the same time, presents the three primary problems of jet propulsion:

1. What fuel do we burn, and where does the oxygen for the combustion come from?
2. How do we obtain the most efficient combustion?

and

3. How do we best utilize the high pressure resulting from the combustion for the propulsion of our body?

Seeking the solution to our first problem, that of the fuel and the oxygen, we will probably get a quick answer: there is a variety of fuels available today, meeting a host of different requirements, but why not take the most widely used, gasoline, or some similar petroleum product? And as far as the oxygen is concerned, there should certainly be enough in the surrounding air, enough, anyhow, to support combustion everywhere else. Well, both answers make sense, but the one concerning the oxygen in the surrounding air holds true only in regions where there is air. What about space? There is no air in space, consequently no oxygen. No oxygen, no combustion.

It appears then, that before we continue our discussion, we must make a clearcut distinction between two different types of jet power plants; those that rely upon the oxygen of the surrounding air to support the combustion, so-called air-breathing engines, and those that carry their own oxygen in some form or another, rocket engines. It is evident that air-breathing engines can breathe air only where there is some, and that means their operating range is limited to the atmosphere. Rocket engines, on the other hand, being self-contained, can operate in a vacuum, in space.

Except for the fundamental fact that both air-breathing engines and rocket engines create hot gases by combustion and then eject these, the two types of powerplants are so different technically that it is more practical and less confusing to deal with them separately.

3. Air-Breathing Engines

Let's begin, then, with the air-breathing engines among which the ramjet holds the greatest promise as a power-plant for supersonic guided missiles, and possibly, manned craft.

The ramjet has been praised, and rightly so, as the simplest engine yet devised. If we disregard accessories such as fuel pumps and the like, it is the only engine known today without any moving parts at all. The simplicity of its design is intriguing, particularly if you look at the original form conceived around 1910 by the Frenchman, Lorin. He called it the "Flying Stovepipe" and visualized a cylindrical pipe of some length (Fig. 1). He said, and we shall follow his reasoning, if such a pipe is moved through the air at some speed, air passes through the pipe at approximately the same speed. All we need to do is maintain combustion in this pipe, thus reduce the density of the air passing through it, and with it increase its velocity as it leaves the pipe. Consequently, we obtain an exhaust velocity greater than the velocity of the air through which our pipe travels, or in other words, we are producing thrust.

Today, almost 50 years after its conception, the ramjet is still in an experimental stage and there is a good reason for it. In order to have the ramjet engine work at its best and fully utilize its potentialities, we must work in regions of high supersonic velocities. As in the case of turbojet-driven high-speed air planes, control problems, fuel economy considerations and aerodynamic drag at these speeds require operation at high altitudes, let's say somewhere between 50,000 and 100,000 feet as optimum for the ramjet.

These altitudes and speeds, of course, were beyond the reach of Mr. Lorin and his contemporaries 50 years ago, and the science of aerodynamics dealing with supersonic flow was no common engineering knowledge at that time either. So, the idea of the "Flying Stovepipe" was conceived long before it could be put to practical use, and as such is a parallel to the airplane and jet propulsion in general.

Let's have a little closer look now at the problems baffling today's engineer concerned with the development of ramjets.

We will assume our flying pipe moving along at a speed of, let's say, 2000 miles per hour, roughly three times the speed of sound, at an altitude of 80,000 feet. The atmospheric pressure at this altitude is sufficiently low to allow us the establishment of a high pressure ratio between combustion chamber pressure and ambient pressure, even if the chamber pressure should not be very high. This high pressure ratio, we remember, is required to obtain our needed high jet exhaust velocities. But how do we go about establishing this high pressure ratio? All we have assumed so far is a low ambient pressure, that is a small denominator in our ratio. We still need a large numerator, a high combustion chamber pressure. Even if combustion alone sufficed to build up this high chamber pressure, we would still be confronted with the problem of initiating this combustion, that is igniting a fuel-air mixture traveling at three times the speed of sound, at an absolute pressure of not quite 1/2 psi. There is no indication that anyone will ever succeed in solving this problem under such adverse conditions. To support combustion, and especially to establish ignition of a combustible mixture, we must go to substantially lower air flow velocities and higher air density and pressure. So, our problem of establishing the proper conditions for our combustion boils down to converting a high-velocity, low-pressure airstream approaching our pipe from the outside, to a low-velocity, high pressure airstream inside our pipe in which we want to establish and maintain combustion; and then go back to a high-velocity airstream of hot gases emerging from our exit opening. At first glance, the problem appears simple and one regularly encountered in fluid dynamics dealing with subsonic flow. As long as our fluid, in this case our air or combustion gases, is traveling at a speed lower than that of sound, we only need to have our duct diverge. By doing that, we slow down our gas flow and increase its static pressure. Vice versa, we accelerate our gases by contracting our duct. This is common engineering knowledge and true as long as we deal with velocities lower than the speed of sound, where we consider gases to be incompressible. But, remember, we are traveling at three times the speed of sound, in other words, three times as fast as pressure disturbances can propagate. Under these conditions, the compressibility of gases can no longer be neglected in our flow considerations, and the picture changes radically.

It is obvious that we cannot digress here into a lengthy discussion of supersonic flow, a very intricate and involved branch of aerodynamics, but I must ask you to put up with my

stating a few salient facts essential to an understanding of the functions and operations of a ramjet engine.

The first problem that confronts us is: How do we go about converting the supersonic flow approaching our engine into the subsonic flow required inside our combustion chamber, and convert back to the supersonic flow of gases emerging from our exhaust opening?

Early experimenters thought that speeds faster than sonic could not be created because they found, and that still holds true, that gases escaping from a chamber through an orifice will never reach, in this orifice, a speed higher than that of sound, regardless of how high the pressure in their chambers are. When the Mach number, called M, is the ratio of actual speed to local sonic speed, it was thought that only Mach numbers smaller than 1 could be realized, but Mach 1 was the limit. But then theoreticians went to work and, by combining a few simple relations known in thermodynamics and flow dynamics, arrived at two basic equations which were to be the foundation of a great deal of important work done subsequently in aerodynamics:

and

$$\frac{d\varrho}{\varrho} = -M^2 \frac{dV}{V} \tag{1}$$

where

$$\frac{dA}{A} = (M^2 - 1)\frac{dV}{V} \tag{2}$$

ϱ = Density

M = Mach number

V = Velocity

A = Area

Equation (1) shows that, as the flow Mach number increases below Mach 1, the velocity of the stream will increase faster than the density decreases; so that in order to fill a passage, its area must decrease. On the other hand, above Mach 1, the density decreases faster than the velocity increases and, hence, the passage must widen to make room for the added volume.

Equation (2), closely related to the first one, yields the following, far-reaching conclusions:

1. If the speed of the gas stream is subsonic, that is M smaller than 1, $M^2 - 1$ is minus, and an increase of area yields a decrease of velocity.

2. If, on the other hand, the speed of the stream is supersonic, or Mach larger than 1, $M^2 - 1$ is plus, and an increase of area yields an increase of velocity;

3. If we have a minimum section in our passage, a so-called throat, $dA = O$, and we can have either $M = 1$, that is sonic speed, or $dV = O$. In other words, sonic speed is obtained only at the smallest section, called "throat".

This means in simple words that a supersonic stream may only be created in a duct by first having a contraction at which the Mach number will be one and only one, and in direct opposition to the behavior of a subsonic stream, a supersonic one has its velocity increased

by a widening of the walls. Thus, we arrive at a passage which first converges and then diverges, a so-called deLaval nozzle, and we see the velocity and pressure pattern of a gas stream flowing through this passage (Fig. 2). As you can see from the two curves shown, this type of nozzle meets both of our requirements. If the approaching flow is supersonic, our nozzle converts it to subsonic; if the approaching flow is subsonic, our nozzle accelerates it to supersonic, provided we have a sufficient pressure ratio. In both cases, we have speed of sound in our narrowest cross-section, the throat.

To repeat, we can use this converging-diverging nozzle to solve both of our previously-stated problems. We can use it at the forward entrance to our pipe in order to convert a supersonic, low-pressure airstream into a subsonic high-pressure airstream which we need in our chamber in order to establish and maintain combustion. And, downstream from our combustion chamber, we use the same type of converging-diverging nozzle to accelerate our low-velocity, high-pressure gas stream back to supersonic velocity required to yield the necessary thrust.

Let's visualize a picture, then, which looks about like (Fig. 3). Here on this picture you see a line called "Normal Shock". What is a normal shock? It is a very thin discontinuity in an air or gas stream in which an abrupt pressure and temperature rise takes place. It is an extremely important phenomenon in supersonic flow because it always occurs in changing from supersonic to subsonic flow. To visualize a normal shock is about as difficult as to visualize electricity, but nevertheless, its effect on gas flow can be fairly well established. Flow through a normal shock is accompanied by a substantial gain in entropy which means that, although air passing through the shock experiences an increase in pressure and temperature, not all the pressure is recovered. In other words, the final pressure after a shock is somewhat lower than it would have been had the flow slowed without the shock. Such a normal shock has the undesirable quality of becoming more intense with increased Mach numbers, and this increased shock intensity is reflected in higher losses.

The bottom curve on Fig. 4 shows us that at higher Mach numbers, the shock losses become so substantial that there is not much left of the pressure recovery we are trying to obtain by slowing down our flow. But this diagram shows more; there are three additional curves resulting from combinations of normal shock with one or more so-called "oblique shocks". And these curves evidently look much more favorable, as far as our possible pressure recovery is concerned. So, they are worth investigating. But now, what is an "oblique shock" and how do we go about establishing one or more to improve our pressure recovery?

The fundamentals of an oblique shock can be explained in comparatively simple terms using Fig. 5 and 6. Consider the wave fronts produced by a very small particle moving at supersonic speed. For instance, let the particle be fired from a gun at a velocity of Mach 2, and let us examine the flow patterns at equally-spaced points A, B, C, D along the path of this particle. By the time the disturbance created by the particle when it was at A has reached a point midway between A and B, the particle itself has reached point B, because it travels at twice the speed of sound. By the time the particle reached C, the wave front from A has reached B, while that from B has reached halfway to C. We see, therefore, that a line drawn from the particle at any instant can be made tangent to the wave fronts and, further, this tangent, being the locus of all the wave fronts from A to D, will, therefore, have the maximum rise or drop in pressure. Consequently, a conical wave front will develop around this particle, with the angle of the cone dependent on the Mach number at which the particle travels. If we replace our particle by a sharply pointed cone or sharp-edged wing tip, the angle of our wave front will spread due to the displacement of space and crowding of air molecules (Fig. 7). Such a wave, inclined to the free airstream at an angle greater than that dictated solely by the Mach number, is called an oblique shock. Such oblique shocks slow down the approach velocity of our air flow, but they are very weak compared to a normal shock and entail much smaller losses. Therefore, we can improve our inlet conditions if we choose an inlet configuration with one or more oblique shocks followed by a normal shock. The oblique shock or shocks will gradually slow down our air flow to lower supersonic Mach numbers without

appreciable entropy gains, and the normal shock, being weaker due to the slower flow approaching it, will also entail a smaller loss. The result is that the more oblique shocks we can put in front of our normal shock, the better our overall compression efficiency, especially at high cruising Mach numbers of our powerplant (Fig. 4). Fig. 8, taken from a recent articie published in the Journal of the American Rocket Society shows a few possibilities of such configurations.

So far, we have succeeded in converting the supersonic flow approaching our engine into a subsonic flow. Immediately behind the normal shock, however, the Mach number of this air stream is still rather close to 1, still too high to establish combustion. But, behind the normal shock, we can revert to conventional subsonic methods to slow down our flow further and all we need to do is gradually widen our duct again, in other words, attach the diverging part of our converging-diverging inlet system. In contrast to the converging inlet, the supersonic diffuser, this diverging passage behind the throat is called the subsonic diffuser; it connects our inlet with the combustion chamber and delivers to it an air flow of reasonably low velocity and high pressure in which, with a few tricks, we can establish our combustion. We must emphasize the gradual widening of our subsonic diffuser, by the way, to prevent flow separation at the walls which would result in a fast core stream and an extremely uneven velocity profile in our duct, a rather unfavorable condition under which to establish and maintain combustion.

All the time we have talked about velocity and pressure, we have completely disregarded another important factor in these considerations, that is the temperature. Since the air entering our ramjet inlet is gradually compressed, it experiences a considerable temperature rise which, of course, is very helpful and desirable to improve our combustion conditions. The temperature and pressure in the combustion chamber depend on the flight velocity, the cruising altitude, and the pressure recovery obtained from the inlet; therefore, they vary over a wide range depending on operation and conditions. Air temperatures between 100 and 1200 F might be encountered and pressures might range from several atmospheres down to a fraction of atmospheric pressure. But we can use our example of an engine traveling at Mach 3 at 80,000 feet altitude. The ambient pressure at this altitude is roughly .4 psi absolute. If we take an efficient inlet configuration, we should be able to obtain a pressure recovery of 20:1, that is, we can deliver a pressure of approximately 8 psi to our combustion chamber; the stagnation effect would raise our air temperature to about 700° F and the air would travel through our combustion chamber at roughly 300 to 500 feet per second, depending on design details.

How do we go about establishing and maintaining combustion under such adverse conditions? First of all, we need a source of ignition; but since we are dealing with continuous combustion, an initial spark is all we need, so some kind of an expendable spark plug or pyrotechnic device will generally suffice. Then we need an injection system by means of which we can distribute our fuel (let's assume it to be a liquid hydrocarbon) over the entire cross-section of our combustion chamber. Here is another problem: we are injecting into a high-velocity airstream which travels, as we said, at 300 to 500 feet per second. Elaborate atomizing nozzles with wide spray patterns are no longer very efficient under these conditions because the spray has no chance to develop and has no penetration. So, we are forced to place a grid of injection nozzles or openings across our combustion chamber inlet to insure good fuel-air distribution in our air stream. And it looks like that's all we need; we have air, even hot air, we have fuel injected through a well distributed system, and we have a spark. But all the tricks that we have used to improve our combustion conditions are still not sufficient. We need to create still lower local air velocities and regions of turbulence to establish and maintain a stable flame front.

These required local air velocities must obviously be lower than the flame propagation velocity; otherwise the flame will tend to escape from the confines of the combustion chamber, a phenomenon known as "blowout". For the purpose of obtaining such low velocities, a so-called "flame-holder" is necessary as a starting point for combustion. It usually consists of

some type of baffle placed across the cross-section of the combustion chamber at some distance downstream from the fuel injection system. The characteristics desirable in this baffle, or whatever it may be, are that it provides a high degree of turbulence coupled with low axial air velocities and a minimum drop in total pressure. Obviously, these three requirements do not go hand-in-hand, and usually the pressure drop will suffer in order to meet the other two.

Although a great deal of the work done on combustor design and test results are still classified for obvious reasons, I can give you a general idea of what such flame-holding devices may look like. Most of them consist of open or closed cones in annular or grid arrangements, or cans, similar to those employed in turbojet engines (Fig. 9). Once the fuel-air mixture is ignited in the wake of these flame holders, the flame is stabilized in this wake due to recirculation. In other words, hot combustion products are fed back into the immediate wake of the baffle where they serve as a continuous source of ignition for the fresh fuel-air mixture flowing past the edge of the flame holder. A solid flame front, attached to the baffle or can will then develop and fill the entire cross-section of the combustion chamber.

Here we must inject an important thought. Since there are no moving parts in the flow of combustion gases in the ramjet engine, in contrast to the gas turbine, we need not dilute the combustion air with secondary air, but, instead, we can operate much more efficiently at higher temperatures (around 3000^{0} F) with fuel-air ratios at or near stoichiometric. That means that all air entering our ramjet participates in the combustion and proper fuel distribution and fuel-air mixing become of paramount importance, if high combustion efficiencies are to be obtained, and they certainly are. The location of the fuel injection elements and the baffle arrangement must then be well matched in order to obtain optimum control of the fuel distribution relative to the flame-holding elements.

The combustion established in the wake of our flame holder will, of course, require a chamber of a certain length for its completion and a thorough mixing process. The longer we make this chamber, the more stay time we give our fuel-air mixture and combustion gases, the more certain we can be of complete reaction of all fuel available. But engine size and weight considerations force us to compromise and put a practical limitation on the length of the combustion chamber, depending essentially on its cross-sectional area and the temperature and pressure conditions peculiar to the design.

In the case of engines operating at high Mach numbers with higher overall temperature levels, cooling of the chamber walls may become necessary, but, in most cases, we can establish favorable conditions of heat transfer to the ambient air and can get along without making special provisions for chamber wall cooling.

At the rear end of our duct, we attach a converging-diverging exit nozzle, as previously discussed. Thus, we accelerate our hot gases and, having a sufficient pressure ratio between combustion chamber and ambient air pressure, we expand the gases supersonically to high exhaust velocities and, thus, give our vehicle the required thrust.

That, essentially completes the description of the ramjet and we obtain a configuration as shown in Fig. 10 as one possibility. As we found out earlier, the cruising Mach number and the altitude at which the engine is to operate will substantially affect various design parameters, and the picture shown here merely gives an example of what such an engine may look like.

Fuels used in ramjets have to meet the same requirements as those employed in turbojet engines, that is, highest heat release per unit volume, ease of handling, ease of combustion and availability. Ease of handling tends to rule out solid and gaseous fuels since, by this criterion, liquid fuels are preferable. Unstable fuels that tend to decompose or become dangerous in storage would also have to be ruled out. Of the stable liquid fuels presently available in reasonably large quantities, petroleum hydrocarbons give the highest heat release per unit volume and are, therefore, generally used for ramjet propulsion work.

One of the main problems encountered in ramjet engine work arises as the result of the rather sensitive tuning and adjustment that such engines require. Inlet configurations, fuel-air mixing ratio, details of the combustor and exit nozzle design must be closely correlated to obtain maximum performance. As we remember, ramjet inlet diffusers are sensitive to velocity changes because the shock pattern is developed for only one cruising Mach number. Poorer pressure recoveries result from deviations from the design velocity and these lower recoveries, of course, affect combustion efficiencies. In a similar manner, combustors are developed for certain conditions in the combustion chamber, and, if we vary these conditions substantially, we experience unstable or inefficient combustion, possibly blowout. Such instabilities or lower combustion efficiencies will necessarily affect our pressure buildup in the combustion chamber and, with it, our exhaust velocity which, in turn, will have an adverse effect on the inlet conditions. This interplay of inlet, combustion and exit makes ramjet development rather complex. But not only that, it puts one of the major limitations on the practical application of the ramjet engine. As you can gather from what was just said, the ramjet is not very flexible in its operating range and is generally expected to function only at or near the design conditions of flight Mach number, altitude and fuel-air mixture ratio. It will not normally be able to take care of acceleration, except within rather narrow limits such as those required for thrust control. I am not saying that ramjets capable of climbing or covering a wide band of velocities cannot be built, but they would have to have adjustable, not only adjustable but self-adjusting, inlets operating efficiently under a variety of conditions of speed and altitude; and along with an adjustable inlet, this type of engine would require an adjustable exit nozzle to account for the varying pressure levels in the combustion chamber.

Therefore, we can safely say that, within the near future, a ramjet operating under greatly varying conditions will be the exception rather than the rule. The simplicity of this engine, however, and its high performance at supersonic speeds in level flight make it an excellent choice for long-range guided missiles or possibly as an auxiliary powerplant for aircraft operating at high altitudes. As you may have gathered from our description of the ramjet engine, it cannot develop any static thrust, such as the rocket or the turbojet. It must be brought to or at least near its operating altitude and velocity before it can function properly. That means that it must be boosted by some other vehicle, such as a rocket or a high-speed airplane. This fact, together with control and internal stability problems and the need for elaborate and expensive ground testing facilities, and the cost of full-scale ground testing is the main drawback of the ramjet engine. But, for sustained propulsion at supersonic speeds, no other engine can compare with it. Its outstanding features are its high power per unit frontal area and per unit weight, its high specific impulse as compared to rockets, and its mechanical simplicity. This mechanical simplicity is deceiving, however, as you may have gathered from the foregoing discussions. Perhaps more than any other engine, the ramjet demands a thorough knowledge of the aerodynamic and combustion processes upon which its successful functioning depends. Scaling from small models to large ones is a combustion, not a mechanical problem. An insight into the individual processes, which can only come from a wealth of supporting research into the fundamentals, is the best hope for extending ramjet sizes and velocity ranges beyond those where laboratory testing facilities on full-scale units become impractical.

The term air-breathing engines includes two more important types of jet powerplants; the turbojet and the pulsejet. Operation and technical details on the subsonic turbojet engine are common engineering knowledge today and need not be discussed here. Advanced types of turbojet engines, however, are now under development and production, which feature supersonic inlet diffusers similar to those discussed for the ramjet, and the door is open for turbojet engines operating in the low supersonic regime. Details are still under security wraps, but the knowledge of such turbojet-driven missiles as "Matador" and "Regulus" indicates definite potentialities of this engine for guided missile propulsion.

The other air-breathing engine which has been used for propelling guided missiles is the pulsejet, a cousin of the ramjet and well-known in the form of its prototype, the German V-1, or buzz bomb. The pulsejet takes over where the ramjet becomes less efficient, at lower

flight velocities in the subsonic or, possibly, low supersonic regime. In contrast to the ramjet which must be boosted to its operating conditions because it depends on the ram pressure produced by high velocities, the pulsejet operates intermittently and develops static thrust.

The pulsejet engine consists of a duct externally resembling the ramjet, but, unlike the latter, it is equipped at its forward end with a grid-type valve made up of a number of shutters (Fig. 11). A fuel-air mixture ignited behind this valve develops a pressure that closes the valve and, in turn, creates a jet effect at the rear end. Because of the momentum of the gases traveling through the pipe, a reduced pressure is effected by which the shutters are opened and a fresh charge of air is admitted through the valve grid. Due to the temperature of the duct and part of the flaming gases backing up into the tail pipe, the new charge is ignited, and the operation proceeds without the benefit of ram air or further operation of the spark plug. The shutter valve operated by a series of pressure changes is the only moving part in this engine.

Once the operation is started, the cycling continues at a frequency determined by the design of the duct and its operating conditions. In order to obtain the greatest efficiency, the explosions of the fuel-air mixture must be in resonance with the compression wave frequency of the whole system, a fact which explains the name "resonant jet" sometimes used for this powerplant.

The pulsation generated by this cycle creates an ear-shattering noise and extremely strong vibrations, so strong, in fact that, in the V-1 buzz bomb, normal fasteners such as rivets, bolts and screws did not stand up, and welded construction had to be used wherever possible. These vibrations also limit the service life of the shutter valve to minutes, or at best, a few hours.

The mechanical design of this engine is by far the simplest and least expensive known for static operation, that is, development of thrust without motion and at low cruising speeds. But present indications are that speeds obtainable with pulse are below those generally desirable for guided missiles with regard to vulnerability. The applicability of pulsejets, at least in their present state of development to guided missiles is, therefore, limited and so must be our discussion of this engine within the subject of this paper.

4. Rocket Engines

This completes our discussion of the air-breathing engines applicable as powerplants for guided missiles. As we remember, these types of engines can operate only within the atmosphere as they draw the oxygen necessary for combustion from the ambient air. In contrast to this, the rocket engine draws its oxygen from a supply source carried aboard the vehicle it propels. A rocket-driven missile, be it a small projectile or, one day, a large space ship, is self-contained; it requires no atmospheric oxygen for its operation; it can function in a vacuum; its thrust is practically independent of its environment. Moreover, it appears to be the simplest means for converting the thermochemical energy developed by a fuel-oxidizer combination into the kinetic energy of a jet of flowing gases. But, on an energy basis, the rocket engine can work efficiently only at extremely high flight speeds, considerably above sonic speed. The enormously high fuel consumptions associated with such high speeds, rule out the rocket engine as a competitor of existing means for propelling aircraft unless extremely high speeds are to be obtained. The rocket engine is a divice which has the potentialities of attaining objectives unattainable with other means of propulsion.

Since our discussion began with air-breathing engines, we may now be able to simplify the explanation of the rocket engine by asking, "Where are the differences?" There are many important points in which the two differ, so many, in fact, that it is simpler to turn the question around and ask, "What do they have in common?" The answer to that is short: both develop hot gases by burning fuel in a combustion chamber and then expand these gases across

an exhaust nozzle. Just about everything else is different. Since we do not feed atmospheric air into the rocket engine, we have no problems of inlet configuration or pressure recovery, nor do we encounter complex investigations of combustion in a fast-flowing airstream at low pressure. But we do have a host of problems elsewhere, as we shall see in a few moments.

Basically, the rocket engine consists of a combination chamber with an exit nozzle, and an appropriate propellent supply. Depending on the requirements to be met by the engine and its purpose, the fuel can be either solid or liquid, and most considerations of rocket engines are based on this classification into solid-propellent and liquid-propellent rocket engines.

Fig. 12 shows a schematic outline of a solid propellent rocket motor. As we see, all the propellent, a substance remotely resembling gunpowder, is stored inside the combustion chamber and, upon ignition, gradually burns away. The fact that the combustion chamber is propellent storage at the same time is of great importance because it determines both the advantages and the disadvantages of this powerplant. Since the combustion chamber is a heated pressure vessel, the thickness of its walls must be comparatively large and there are practical limits to its size due to weight consideration. Consequently, solid propellent rockets generally have rather short burning times, generally not exceeding 1/2 minute. These engines cannot be controlled or governed once the charge is inserted and ignited, nor can they be turned on and off at will. But these same features that limit the suitability of solid-propellent motors for guided missile propulsion, make it very desirable for other purposes where a short burst of power is the main requirement, such as for jet-assisted takeoff for aircraft, rocket projectiles for meterorological sounding purposes and as boosters for other missiles. In these cases, the simplicity of the solid propellent motor makes it far superior to the liquid propellent powerplants with all the tanks, pumps, valves and lines necessary to feed the propellents into the combustion chamber.

But, where sustained combustion over a longer period of time is required, such as in long-range guided missiles and, particularly, in future space ships, liquid propellent rockets are the answer as their burning time is limited only by the size of the propellent tanks, but not by the size of the engine itself. Such liquid propellent engines were used in the German V-2 which developed approximately one-third horsepower and they propel such missiles as the Wac Corporal, the Viking, the Aerobee and the well-publicized Nike anti-aircraft rocket, also by the way, the fastest airplanes ever built, the Bell XS-1 and the Douglas Skyray. In the liquid-propellent engine, we have again the combustion chamber with the exhaust nozzle, but the charge is replaced by a propellent feed and injection system as shown in Fig. 13.

Fuel and oxidizer, the two propellents, can be either pressure fed, employing a high pressure inert gas for pressurization (Fig. 14), or they can be pump fed, the method depending on design and operating conditions, especially on the total amount to be fed into the engine. In the case of high-powered engines of substantial burning times, the pressure-feed system would demand excessive tank weights and become impractical while, in smaller powerplants, its simplicity may be the over-riding viewpoint. But, there is every indication that for long range rocket missiles, and eventually for spaceships, the compact and lightweight pump feeding system is the solution. Let's not underestimate the problems involved, though: in the V-2 engine, for instance, which developed about 55,000 pounds of thrust or roughly 320,000 horsepower, a total of close to 300 pounds of propellents were fed into the combustion chamber every second, roughly one drum of alcohol and a drum of liquid oxygen per second. This tremendous consumption, naturally calls for substantial power requirements on the pumps, in the case of the V-2, 600 horsepower. Even though this power had to be available only for 60 seconds of powered flight of the missile, the weight penalty would have been very tangible, had an electric motor-driven pump with the necessary batteries been used. Instead, a solution was developed which is still considered the most elegant for this purpose, from the standpoint of both weight and size. In the V-2, both propellent pump impellers were mounted on the shaft of a gas turbine which was driven by steam developed by reacting concentrated hydrogen peroxide with calcium permanganate as catalyst. These two chemicals were pressure-fed to a gas generator as shown in Fig. 15.

Many rocket developers in this country followed this principle and incorporated similar devices into powerplants developed subsequently. Other systems draw hot combustion gases from the chamber, cool them down with some diluent and feed the resulting mixture into the turbine.

The picture you see here is an oversimplification if I ever saw one. A powerplant such as that of the V-2 is a maze of valves, pressure regulators, loaders, pressure switches, sensing elements, relays, tubes, pipes and wires. A great deal of these are needed to insure proper ignition and a smooth start of the engine. If you don't see what I mean, try to dump a drum of alcohol into a container and light it, and watch what happens. The enormous energy releases involved here make it mandatory that the operation of each valve and each relay be timed to split-second operation and, also, that the sequencing of these functions be automatic. In addition to that, safety devices are generally provided which may stop the sequence of events in case of malfunction of important components. This is like having a starter in your car which will not crank up the engine unless the gas tank is filled, the oil level in the crankcase is satisfactory, there is enough water in the radiator, the voltage level in the battery is OK, the ignition coil is in good shape, the points and spark plugs are clean and all the bolts and screws in the car are tight. In short, a smooth start up is among the most perplexing problems in the development of liquid propellent rocket engines.

Ignition, incidentally, is generally provided by a pyrotechnic device, or, if frequent reignition is required, by a small pilot rocket in the combustion chamber. This pilot rocket, in turn, can be ignited by a spark plug.

As fuel and oxidizer are fed into the combustion chamber, they must, of course, be properly atomized and mixed by an injection system. To insure optimum distribution and mixing of fuel and oxidizer, a multitude of systems has been investigated. In the V-2 powerplant, the dome of the rocket motor was equipped with 18 individual injection cups with impinging nozzle systems. A variety of other simpler designs has emerged in the meantime, but many of them are still under security wraps because the question of proper injection and mixing is decisive for the solution of problems of combustion instability still frequently observed in experimental rocket engines. Such instabilities, by the way, can excite extremely strong vibrations leading to gradual or explosion-like destruction of combustion chambers. Improper injector design can also cause locally uneven mixing ratios resulting in chamber burnouts by welding torch effect. We must not forget that, unlike ramjet, pulsejet, and turbojet which use greatly diluted atmospheric oxygen, the rocket employs highly concentrated oxidizers and, consequently, the temperatures obtained in rocket engines are substantially above those reached in air-breathing engines, generally in the order of $5000^{\circ}F$, at combustion chamber pressures of from 300 to 500 psi. Both the temperature attained by the combustion and the composition of the reaction products are influenced to some extent by the oxidizer-fuel mixture ratio and by a phenomenon known as dissociation which tends to lower the peak combustion temperature. This dissociation which is actually a rather intricate physico-chemical phenomenon, may be explained in simple terms as follows: the chemical reaction between a hydrocarbon fuel and oxygen liberates a quantity of heat and forms, in general, carbon dioxide, carbon monoxide, and water. If the heat liberated raises the temperature sufficiently high, these combustion products tend to break up further in heat-absorbing reactions, thus preventing the peak temperature from being as high as it would otherwise have been. The amount of dissociation occurring will depend mainly upon the prevailing temperature and pressure in the combustion chamber. In general, it can be said that the higher the temperature, the greater the dissociation, and the higher the pressure, the less the dissociation.

From a practical standpoint, however, any phenomenon which aids in limiting the peak temperature in the rocket motor without affecting the performance too adversely is beneficial because of the limitations imposed by the materials and cooling methods currently available for the construction of combustion chambers. Even with dissociation, the problem of materials and cooling are formidable and of such nature that they are not generally encountered in other applications. Materials for combustion chamber walls, for instance, must be of heat-

an exhaust nozzle. Just about everything else is different. Since we do not feed atmospheric air into the rocket engine, we have no problems of inlet configuration or pressure recovery, nor do we encounter complex investigations of combustion in a fast-flowing airstream at low pressure. But we do have a host of problems elsewhere, as we shall see in a few moments.

Basically, the rocket engine consists of a combination chamber with an exit nozzle, and an appropriate propellent supply. Depending on the requirements to be met by the engine and its purpose, the fuel can be either solid or liquid, and most considerations of rocket engines are based on this classification into solid-propellent and liquid-propellent rocket engines.

Fig. 12 shows a schematic outline of a solid propellent rocket motor. As we see, all the propellent, a substance remotely resembling gunpowder, is stored inside the combustion chamber and, upon ignition, gradually burns away. The fact that the combustion chamber is propellent storage at the same time is of great importance because it determines both the advantages and the disadvantages of this powerplant. Since the combustion chamber is a heated pressure vessel, the thickness of its walls must be comparatively large and there are practical limits to its size due to weight consideration. Consequently, solid propellent rockets generally have rather short burning times, generally not exceeding 1/2 minute. These engines cannot be controlled or governed once the charge is inserted and ignited, nor can they be turned on and off at will. But these same features that limit the suitability of solid-propellent motors for guided missile propulsion, make it very desirable for other purposes where a short burst of power is the main requirement, such as for jet-assisted takeoff for aircraft, rocket projectiles for meterorological sounding purposes and as boosters for other missiles. In these cases, the simplicity of the solid propellent motor makes it far superior to the liquid propellent powerplants with all the tanks, pumps, valves and lines necessary to feed the propellents into the combustion chamber.

But, where sustained combustion over a longer period of time is required, such as in long-range guided missiles and, particularly, in future space ships, liquid propellent rockets are the answer as their burning time is limited only by the size of the propellent tanks, but not by the size of the engine itself. Such liquid propellent engines were used in the German V-2 which developed approximately one-third horsepower and they propel such missiles as the Wac Corporal, the Viking, the Aerobee and the well-publicized Nike anti-aircraft rocket, also by the way, the fastest airplanes ever built, the Bell XS-1 and the Douglas Skyray. In the liquid-propellent engine, we have again the combustion chamber with the exhaust nozzle, but the charge is replaced by a propellent feed and injection system as shown in Fig. 13.

Fuel and oxidizer, the two propellents, can be either pressure fed, employing a high pressure inert gas for pressurization (Fig. 14), or they can be pump fed, the method depending on design and operating conditions, especially on the total amount to be fed into the engine. In the case of high-powered engines of substantial burning times, the pressure-feed system would demand excessive tank weights and become impractical while, in smaller powerplants, its simplicity may be the over-riding viewpoint. But, there is every indication that for long range rocket missiles, and eventually for spaceships, the compact and lightweight pump feeding system is the solution. Let's not underestimate the problems involved, though: in the V-2 engine, for instance, which developed about 55,000 pounds of thrust or roughly 320,000 horsepower, a total of close to 300 pounds of propellents were fed into the combustion chamber every second, roughly one drum of alcohol and a drum of liquid oxygen per second. This tremendous consumption, naturally calls for substantial power requirements on the pumps, in the case of the V-2, 600 horsepower. Even though this power had to be available only for 60 seconds of powered flight of the missile, the weight penalty would have been very tangible, had an electric motor-driven pump with the necessary batteries been used. Instead, a solution was developed which is still considered the most elegant for this purpose, from the standpoint of both weight and size. In the V-2, both propellent pump impellers were mounted on the shaft of a gas turbine which was driven by steam developed by reacting concentrated hydrogen peroxide with calcium permanganate as catalyst. These two chemicals were pressure-fed to a gas generator as shown in Fig. 15.

Many rocket developers in this country followed this principle and incorporated similar devices into powerplants developed subsequently. Other systems draw hot combustion gases from the chamber, cool them down with some diluent and feed the resulting mixture into the turbine.

The picture you see here is an oversimplification if I ever saw one. A powerplant such as that of the V-2 is a maze of valves, pressure regulators, loaders, pressure switches, sensing elements, relays, tubes, pipes and wires. A great deal of these are needed to insure proper ignition and a smooth start of the engine. If you don't see what I mean, try to dump a drum of alcohol into a container and light it, and watch what happens. The enormous energy releases involved here make it mandatory that the operation of each valve and each relay be timed to split-second operation and, also, that the sequencing of these functions be automatic. In addition to that, safety devices are generally provided which may stop the sequence of events in case of malfunction of important components. This is like having a starter in your car which will not crank up the engine unless the gas tank is filled, the oil level in the crankcase is satisfactory, there is enough water in the radiator, the voltage level in the battery is OK, the ignition coil is in good shape, the points and spark plugs are clean and all the bolts and screws in the car are tight. In short, a smooth start up is among the most perplexing problems in the development of liquid propellent rocket engines.

Ignition, incidentally, is generally provided by a pyrotechnic device, or, if frequent reignition is required, by a small pilot rocket in the combustion chamber. This pilot rocket, in turn, can be ignited by a spark plug.

As fuel and oxidizer are fed into the combustion chamber, they must, of course, be properly atomized and mixed by an injection system. To insure optimum distribution and mixing of fuel and oxidizer, a multitude of systems has been investigated. In the V-2 powerplant, the dome of the rocket motor was equipped with 18 individual injection cups with impinging nozzle systems. A variety of other simpler designs has emerged in the meantime, but many of them are still under security wraps because the question of proper injection and mixing is decisive for the solution of problems of combustion instability still frequently observed in experimental rocket engines. Such instabilities, by the way, can excite extremely strong vibrations leading to gradual or explosion-like destruction of combustion chambers. Improper injector design can also cause locally uneven mixing ratios resulting in chamber burnouts by welding torch effect. We must not forget that, unlike ramjet, pulsejet, and turbojet which use greatly diluted atmospheric oxygen, the rocket employs highly concentrated oxidizers and, consequently, the temperatures obtained in rocket engines are substantially above those reached in air-breathing engines, generally in the order of 5000°F, at combustion chamber pressures of from 300 to 500 psi. Both the temperature attained by the combustion and the composition of the reaction products are influenced to some extent by the oxidizer-fuel mixture ratio and by a phenomenon known as dissociation which tends to lower the peak combustion temperature. This dissociation which is actually a rather intricate physico-chemical phenomenon, may be explained in simple terms as follows: the chemical reaction between a hydrocarbon fuel and oxygen liberates a quantity of heat and forms, in general, carbon dioxide, carbon monoxide, and water. If the heat liberated raises the temperature sufficiently high, these combustion products tend to break up further in heat-absorbing reactions, thus preventing the peak temperature from being as high as it would otherwise have been. The amount of dissociation occurring will depend mainly upon the prevailing temperature and pressure in the combustion chamber. In general, it can be said that the higher the temperature, the greater the dissociation, and the higher the pressure, the less the dissociation.

From a practical standpoint, however, any phenomenon which aids in limiting the peak temperature in the rocket motor without affecting the performance too adversely is beneficial because of the limitations imposed by the materials and cooling methods currently available for the construction of combustion chambers. Even with dissociation, the problem of materials and cooling are formidable and of such nature that they are not generally encountered in other applications. Materials for combustion chamber walls, for instance, must be of heat-

resistant alloy of high heat conductivity; but they are exposed to extreme heat conditions only for minutes and not for weeks, months, or years as in most other applications in industry. Such short service lives allow us, in some instances, to use less expensive material whose creep strength and fatigue properties would not permit their use in long service life applications. On the other hand, such materials are subject to extreme thermal and mechanical shock, which is one of the reasons, by the way, why the use of ceramics for combustion chamber linings is not as widespread as might be expected. Actually, an entirely new branch of metallurgy and material-testing has emerged dealing with extreme heat and shock conditions for thort service lives.

Cooling of rocket motors is generally done by regenerative cooling, that is, by feeding one of the propellents, normally the fuel, through a cooling jacket surrounding the combustion chamber before injecting it; this process has the additional advantage of heating up and, possibly, vaporizing the propellent and thus improving its combustibility. Or, it is done by film cooling which appears to offer the ultimate in permitting high temperatures to prevail in the combustion chamber. It is based on forming, over the inside walls of the chamber and nozzle, a complete film of liquid which, by its evaporation, keeps these surfaces cool. The effectiveness of a combination of regenerative and film cooling was well demonstrated in the V-2 powerplant in which temperatures of 5000°F could be withstood without damage in a motor having a thin carbon-steel inner wall. In some cases, the motor walls and nozzle can be constructed of material heavy enough to absorb the heat during the operating period and satisfactory motors of this type have been built for durations up to approximately 1/2 minute. But, these motors weigh more than others of equal thrust output that employ regenerative cooling, therefore, they are employed only where utter simplicity is desired at the expense of weight. Generally, a combination of regenerative and film cooling is employed and a large number of successful rocket motors of this type have been built.

The performance of a rocket powerplant is most frequently expressed in seconds and is called the "specific impulse". To express performance or effectiveness in seconds may sound strange, but this dimension is quickly explained by the fact that the specific impulse is nothing but the thrust (in pounds) developed per pound of propellent per second. This specific impulse expresses the performance of the powerplant as such, but its value is decisively influenced by the effective heat release of the propellents, so much so, that the term "specific impulse" is frequently applied to weigh the advantages and disadvantages of propellent combinations only, without reference to any particular powerplant. The specific impulse, as we said, is the thrust developed per pound of propellent per second; consequently, its reciprocal is the specific propellent consumption, the required flow rate of propellents to produce 1 lb/sec impulse. For conventional propellent combinations, specific impulses are in the order of magnitude of 210 sec., the specific fuel consumption, therefore, approximately 0.0048 pounds of propellents per 1 lb-sec. thrust. From these figures, we can easily derive the gas exhaust velocity which is specific impulse times gravity, in this case 210 times 32.3, that is, 6,800 feet/second or 4,600 miles per hour.

A short while ago, we mentioned dissociation, the decomposition of combustion products at high temperatures. Because of this phenomenon, not all the thermochemical energy of the propellents is available for transformation into kinetic energy. Consequently, the effective exhaust velocity attained will always be lower than the theoretical value based on the thermochemical energy content of the propellents. In addition to this unavoidable loss from dissociation, there are other losses due to such effects as radiation, fluid friction in pipes and exit nozzle, and incomplete combustion caused by inadequate mixing of some of the propellents.

We have talked quite a bit about how the propellents are fed into the motor and what happens there, but the question still remains, what propellents do we use in liquid rocket engines, what fuels and what oxidizers? Monopropellents in which fuel and oxidizer are combined in one substance are successfully used in solid-propellent rockets, but their introduction into the liquid rocket field has met with considerable obstacles because many of these substances are

inherently unstable. Today's liquid rocket engines use almost exclusively bipropellents, one fuel and one oxidizer, which may or may not ignite spontaneously upon getting in contact with one another. Since our performance, the specific impulse, is essentially dictated by the effective heat release due to the reaction of the two propellents, we should almost be able to obtain the optimum propellent combination by looking up a few figures in physico-chemical tables which give us the heat release of various chemical reactions as computed or determined in laboratory tests. And that is actually so; we can easily determine the propellent combination giving us the highest heat release per unit weight and thus the highest exhaust velocity, and we find it to be oxygen or ozone, and hydrogen. Immediately, we see that there are other problems involved, such as handling. Liquid oxygen is trouble enough with its low temperature and consequent high rate of vaporization. But liquid hydrogen boils at -426°F, not far from absolute zero, and the resulting handling problems are many times more severe. Secondly, we find that liquid hydrogen has a density of only about one-seventh that of water; and that means that, even though the heat release of an oxygen-hydrogen mixture per unit weight may look promising, as expressed by the specific impulse, the same consideration with regard to the unit volume looks quite different. As a matter of fact, practically all of the high specific impulse is nullified by the low density; this is reflected in the fact that the oxygen-hydrogen combination has, at the same time, the highest specific impulse and the lowest "density specific impulse" (specific impulse x specific gravity) of all important propellent combinations.

This example was intended to illustrate the problems involved in choosing an optimum propellent combination for a given design and purpose. Specific impulse, specific gravity, ease of handling and storage, chemical reactivity (that is, ease of combustion and ignitability), availability and, last but not least, the price are to be considered in combinations of such oxidizers as liquid oxygen, nitric acid, hydrogen peroxide, nitorgen tetroxide, and such fuels as ethyl or methylalcohol, gasoline, hydrazine, aniline, furfuryl alcohol, liquid ammonia and a host of others. Although most of today's liquid rocket powerplants such as those of the V-2, Viking, Aerobee, and the Bell XS-1 use liquid oxygen and alcohol, this is not necessarily the last word and continuous research is being carried on to investigate new propellent combinations, or to improve the known ones. This research may well reveal new means of further improving the performance of rocket engines. At this time, we should mention the potentialities of atomic energy for rocket propulsion, especially space travel with its inherently high energy requirements. Often rocket engineers hear statements to the effect that it appears senseless to carry huge quantities of fuel and oxidizer to propel space ships, if a teacupful of fissionable matter and a small atomic reactor on board the ship can do the same trick.

Well, the problem is not quite that simple. Fissionable materials have energy contents many times those of the best chemical fuels, to be sure, and atomic reactors can release enormous quantities of heat, but we still need a substance to be heated up and ejected from our ship to give us the desired propulsive force, we need a so-called working fluid, large quantities of it, and this working fluid has to be carried along in the same manner as propellents for chemical propulsion. Also, we know of the deadly and destructive radiation emanating from atomic reactors which would necessitate extensive and very heavy shielding of these heat sources on board the ship. The weight penalty of such arrangements is obvious and nullifies some of the advantages that atomic propulsion seems to offer. Before atomic energy can be successfully applied to rocket propulsion, another problem is yet to be solved. As we know, the exhaust velocities which can be obtained depend on the temperature to which the working fluid can be raised. This, in turn, is limited by the temperature at which the reactor can be maintained intact, and that, at the present state of development, appears to be the melting point of the fissionable material, in the order of magnitude of 2000°F. Compared to this, we remember that, in chemical propellent motors, we reach 5000°F. Evidently, several problems remain to be solved here but then, atomic energy as a continuous power source has only been explored rather recently, and further sweeping progress is yet to be expected in this field. So, it may not be too long before this new energy source can be successfully applied to rocket propulsion.

A word about power. As we remember, the V-2 powerplant developed 55,000 pounds of thrust, or 320,000 horsepower, roughly twice the power developed by the battleship Missouri. But, recently we have heard an announcement by one of the major rocket engine developers in the country, that there are rocket engines under development which produce several times that thrust. How high can you get? Is there a limit to the size and power developed by a rocket motor? Theoretically, no. But there are certain practical limits; huge combustion chambers would require considerable wall thicknesses which, in turn, would adversely affect the heat transfer between the inside of the chamber and the cooling agent. Gigantic test facilities would be required for ground testing such engines, and the cost of development would be almost prohibitive. Instead, several smaller engines can be arranged in clusters, and especially in cases of extreme power requirements of many million horsepower such as those of future space ships carrying large loads over long distances at extremely high speeds, clustering will in all likelihood be preferred. Another advantage of the cluster arrangement is based on the fact that rocket motors become rather inefficient when throttle action is desired. In cluster arrangements, however, a reduction of the overall power output can easily be effected by turning individual engines off. In other words, clustered or multiple combustion chambers can be operated individually or collectively. Simplicity of repair or replacement and weight considerations also favor the multiple chamber arrangement in many instances.

In conclusion, let me summarize and compare the various jet powerplants discussed, especially with respect to their use in guided missiles and space travel.

The junior partner in this family is the solid-propellent rocket which, because of its inherently limited burning time is used in all cases where short bursts of power are needed. These motors have found application in small artillery rockets, armor-piercing projectiles where high penetration and, therefore, high velocity is required, in small meteorological sounding rockets, and as boosters for missiles from which they are jettisoned, having exhausted their propellents. Recent developments indicate, however, that solid-propellent rocket motors of larger size and for longer burning times are coming into the picture.

The pulsejet, operating in the subsonic region, will be limited in its use for guided missiles due to the vulnerability to high-speed aircraft interception.

The turbojet has recently found its way into the guided missile field after ways were found to operate it supersonically. Present indications are that turbojet-driven missiles will not be capable of exceeding Mach 2 (approximately 1,500 miles per hour); but within this velocity range, the turbojet has distinct advantages over the ramjet in that it is more flexible and its development further advanced.

The ramjet ideally lends itself to the propulsion of missiles covering long distances in horizontal flight and rather high supersonic velocities, and for such long range missiles, is by far the most economical means of propulsion. The chances of intercepting a ramjet propelled missile are very slim today due to its high speed. But, as we remember, ramjet powerplants must be brought rather close to their operating speed and altitude. So they will be either the second stage of a rocket-ramjet missile of which the rocket booster is jettisoned after its exhaustion; or they can be released from fast, high-flying aircraft.

Finally, the liquid-propellent rocket, and we will include here the possibility of atomic propulsion, offers us the only solution presently known to the non-interceptable long range missile, non-interceptable because of its extremely high speeds. And it opens the possibility of space travel because, if a rocket can attain a velocity of 25,000 miles per hour or seven miles per second, the so-called escape velocity, it is possible for its momentum to carry it out of the gravitational field of the earth. The most promising and generally proposed method of attaining this escape velocity is the stepped or stage rocket, a missile that is composed of several progressively smaller rockets, called stages. If the first stage, the largest, is burned out it is jettisoned and the next stage is fired with an initial velocity equal to the velocity attained by the first stage, and so on. Present preliminary blueprints of space ships

call for three such stages with the third finally attaining a sufficiently high velocity to overcome the earth's gravitation and proceed into space.

The multi-stage principle has been successfully applied in many small and medium-sized missiles, but its most spectacular application was in the so-called Bumper missile fired by U. S. Army Ordnance personnel at the White Sands Proving Grounds in February of 1949. This missile consisted of a German V-2 whose warhead was replaced by a Wac Corporal rocket. When the two-stage missile had reached maximum speed, the Wac Corporal was electronically launched, left the nose of the V-2 and gained new speed in addition to the speed of the V-2 that it already had. It reached a maximum speed of 5,000 miles per hour and soared to an altitude of 250 miles, almost twice the height ever reported for any single-stage missile, and definitely outside of the gaseous atmosphere of the earth.

There is no doubt in anyone's mind that a staggering amount of work in research and engineering remains to be done before space travel will become a reality. And, as you well know, engineering and research cost money, a great deal of money for a task of this magnitude.

We can be certain of one fact, however, we are gradually emerging from the fictitious Buck Rogers phase of space travel and progressing toward something more tangible. Every day research of engineering work will bring us closer to the realization of one of mankind's greatest dreams, Junior's dream to become a space ranger one day, and our dream to get away from this war-torn and divided planet of ours, so that we may breathe more freely - under a space helmut!

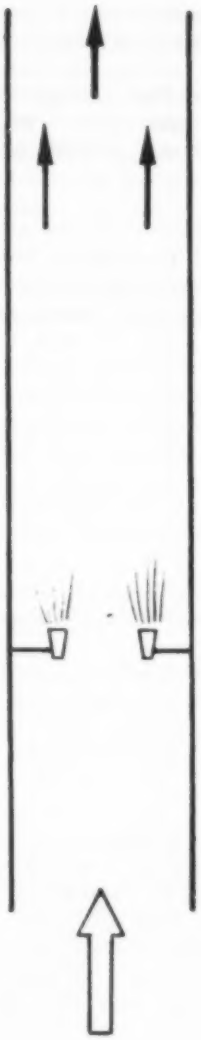
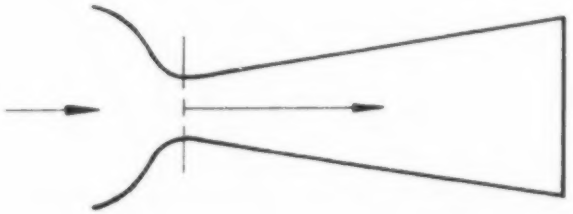
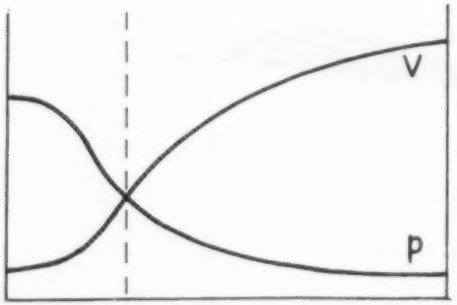


Figure 1





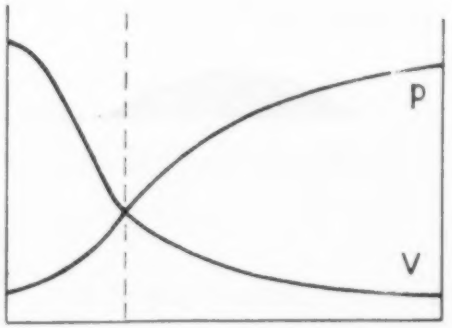


Figure 2

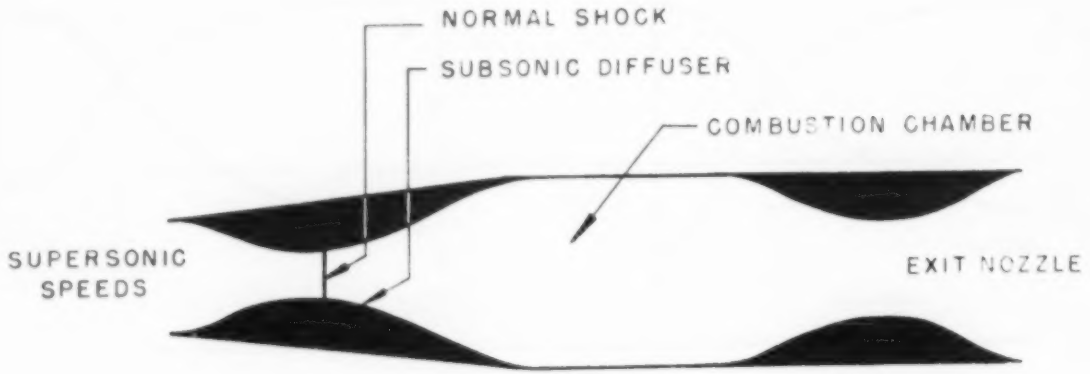

NORMAL SHOCK
SUBSONIC DIFFUSER
COMBUSTION CHAMBER
SUPERSONIC SPEEDS
EXIT NOZZLE

Figure 3

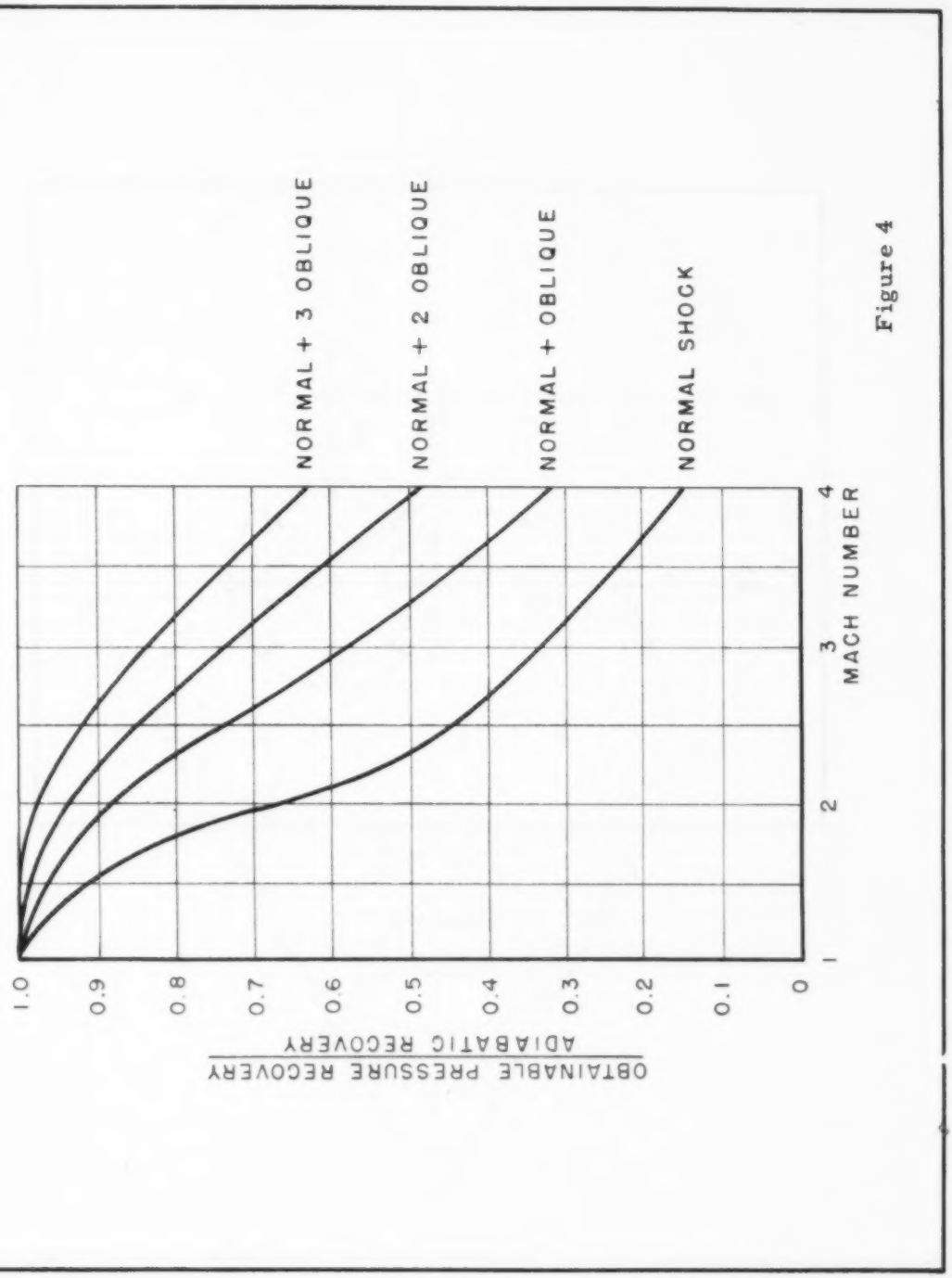

NORMAL + 3 OBLIQUE
NORMAL + 2 OBLIQUE
NORMAL + OBLIQUE
NORMAL SHOCK
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
MACH NUMBER
OBTAINABLE PRESSURE RECOVERY
ADIABATIC RECOVERY

Figure 4

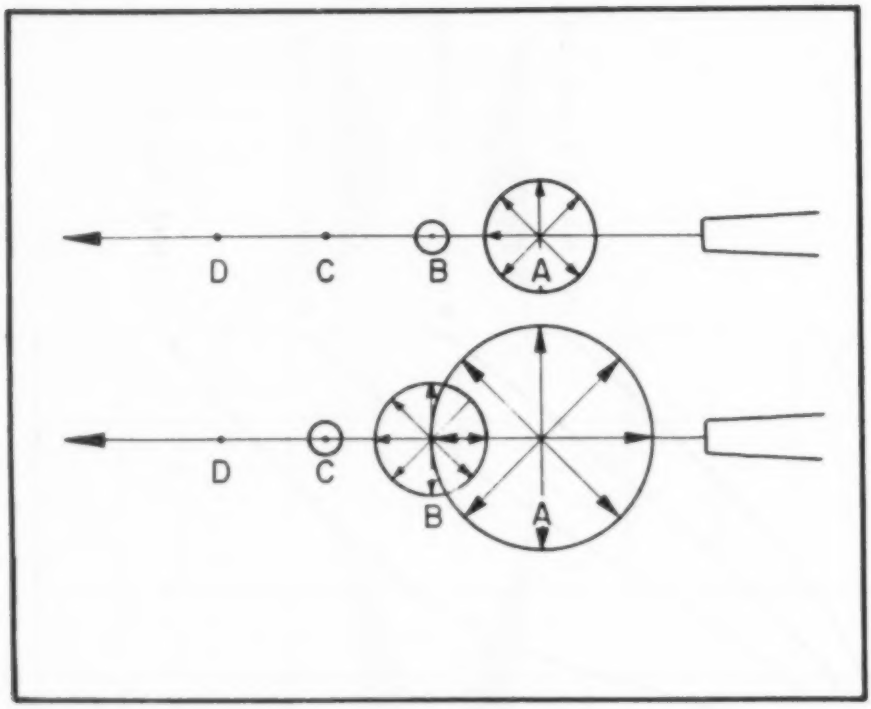


Figure 5

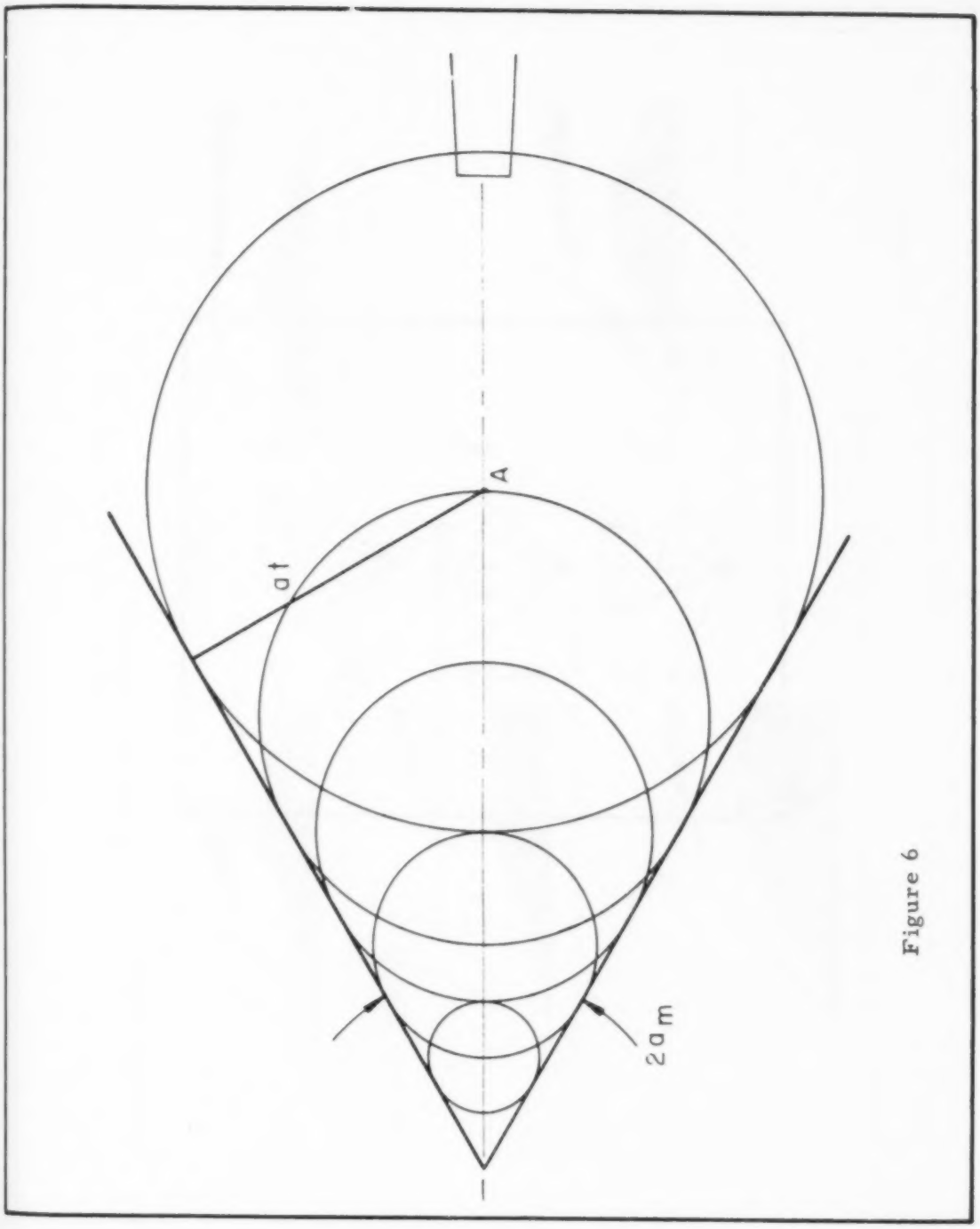


Figure 6

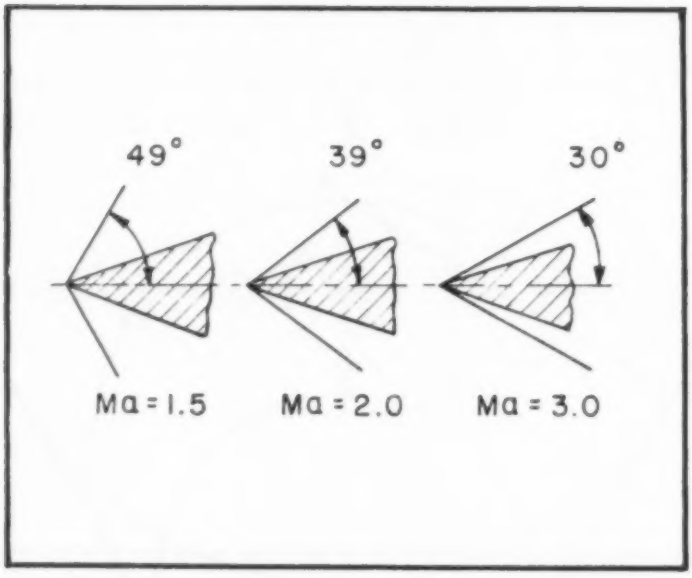
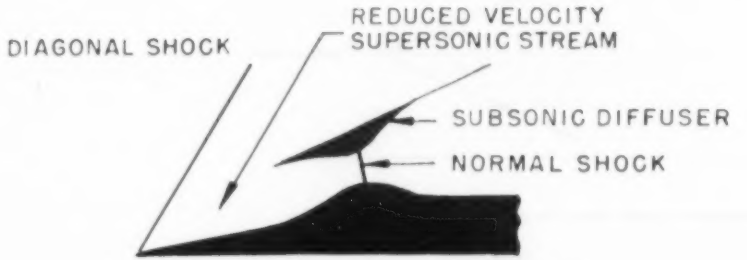
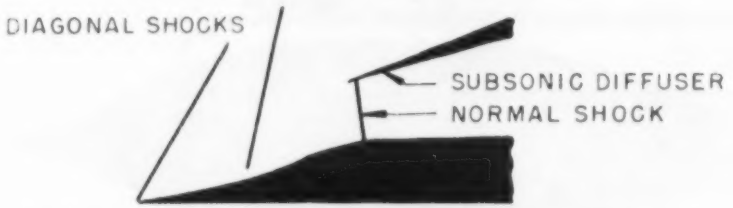


Figure 7

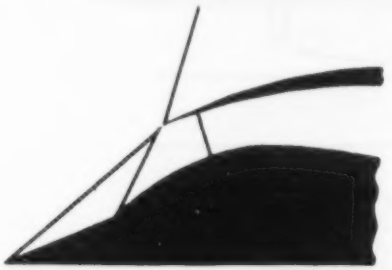


DIFFUSER USING DIAGONAL SHOCK

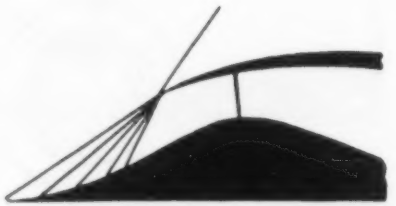
DIFFUSER USING REFLECTED DIAGONAL SHOCK



DIFFUSER USING MULTIPLE DIAGONAL SHOCKS



DIAGONAL SHOCK ENTRY WITH CURVED DIFFUSER



OSWATITSCH DIFFUSER USING COMPRESSION WAVES

Figure 8

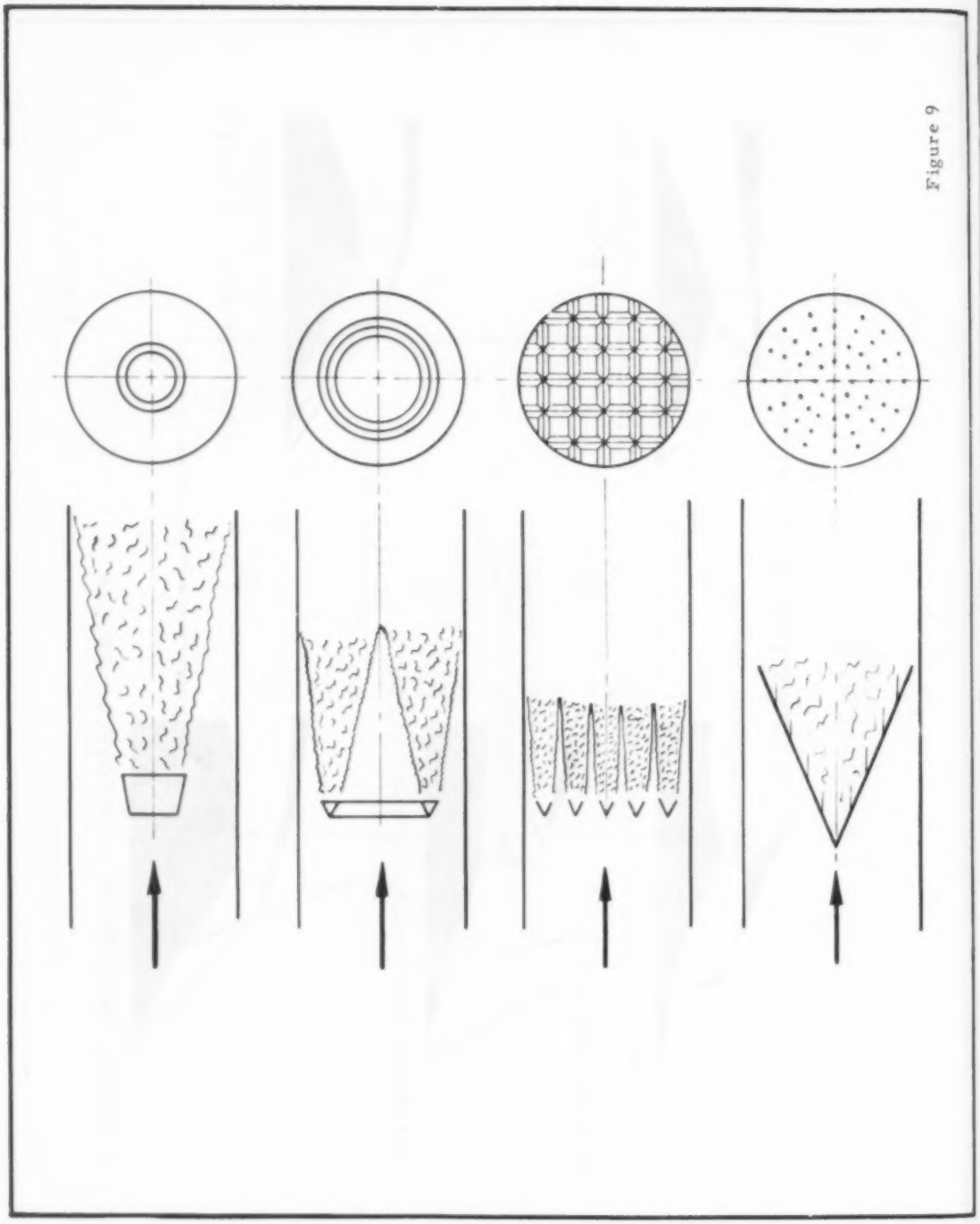


Figure 9

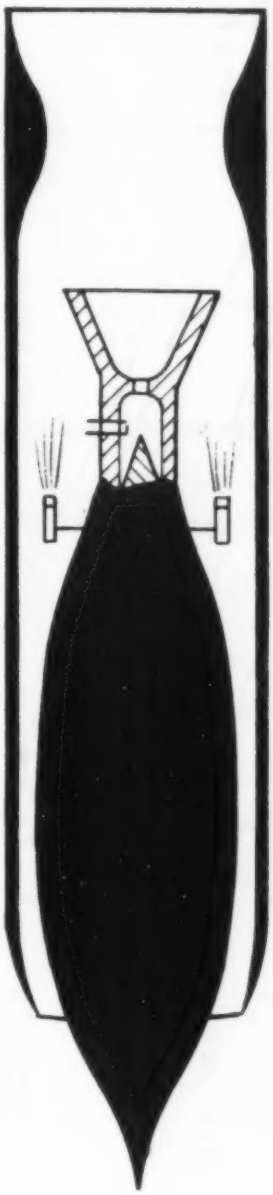


Figure 10

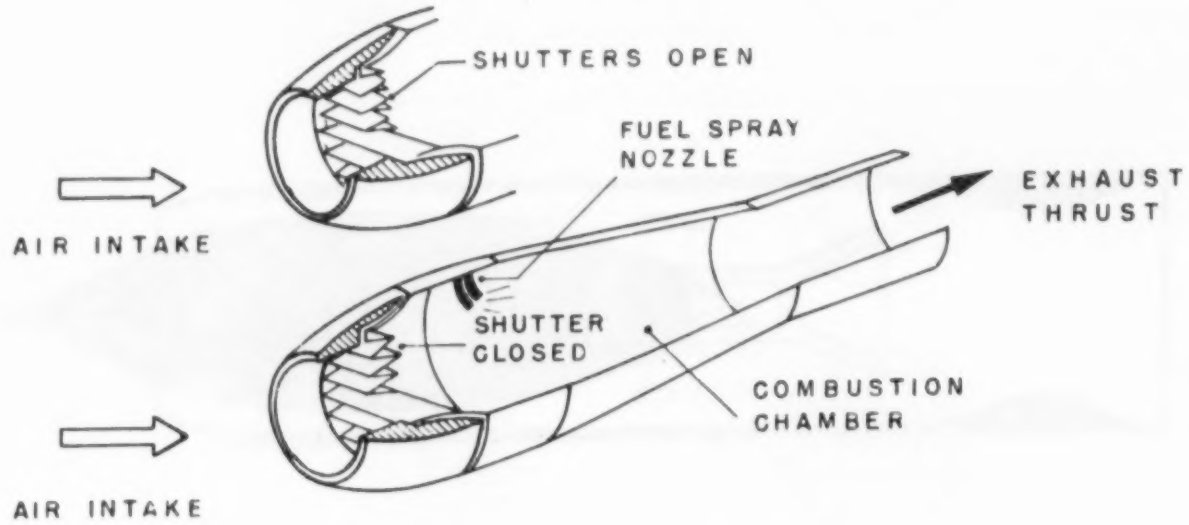

SHUTTERS OPEN
FUEL SPRAY NOZZLE
EXHAUST THRUST
AIR INTAKE
SHUTTER CLOSED
COMBUSTION CHAMBER
AIR INTAKE

Figure 11

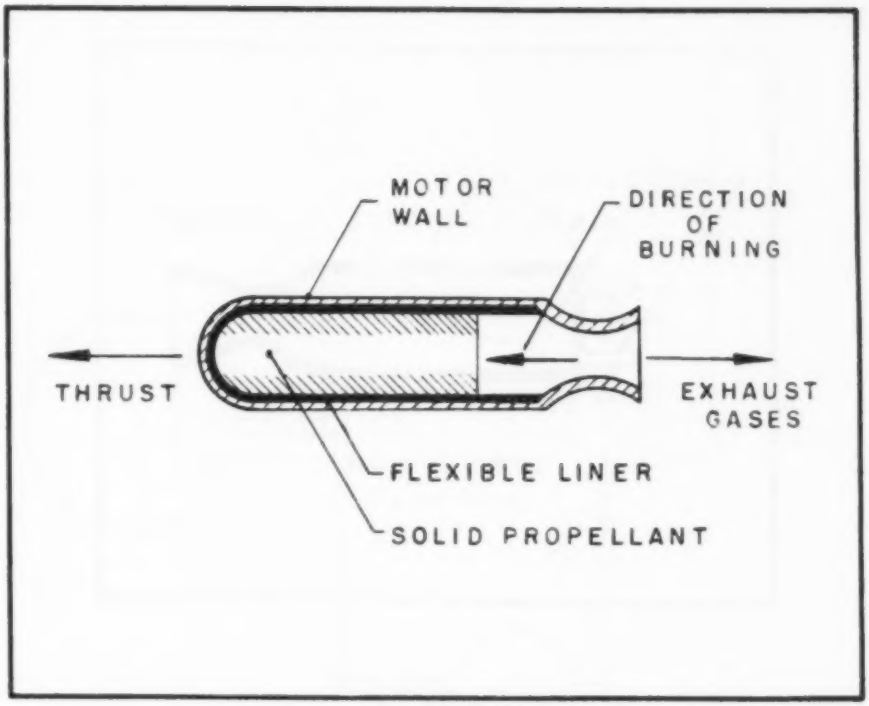

MOTOR
WALL
DIRECTION
OF
BURNING
THRUST
EXHAUST
GASES
FLEXIBLE LINER
SOLID PROPELLANT

Figure 12

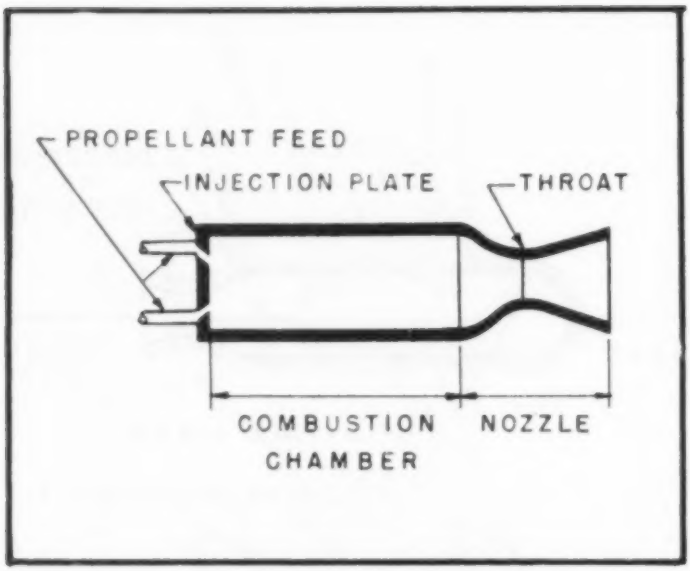

PROPELLANT FEED
INJECTION PLATE
THROAT
COMBUSTION CHAMBER
NOZZLE

Figure 13

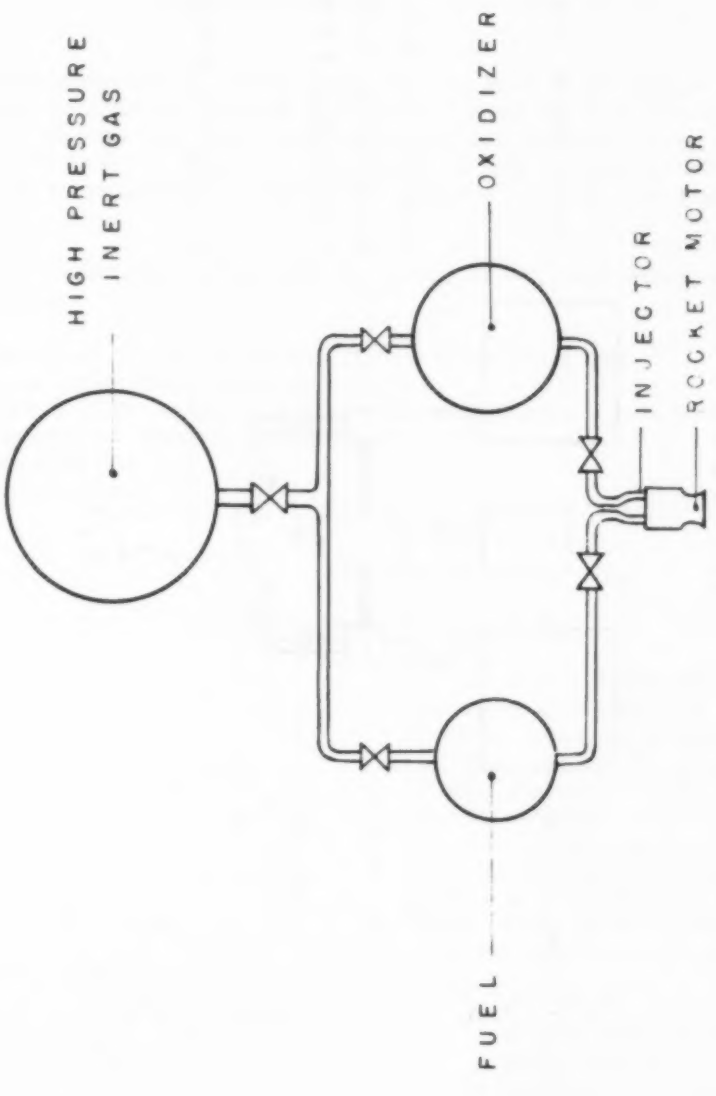

HIGH PRESSURE
INERT GAS
OXIDIZER
INJECTOR
ROCKET MOTOR
FUEL

Figure 14

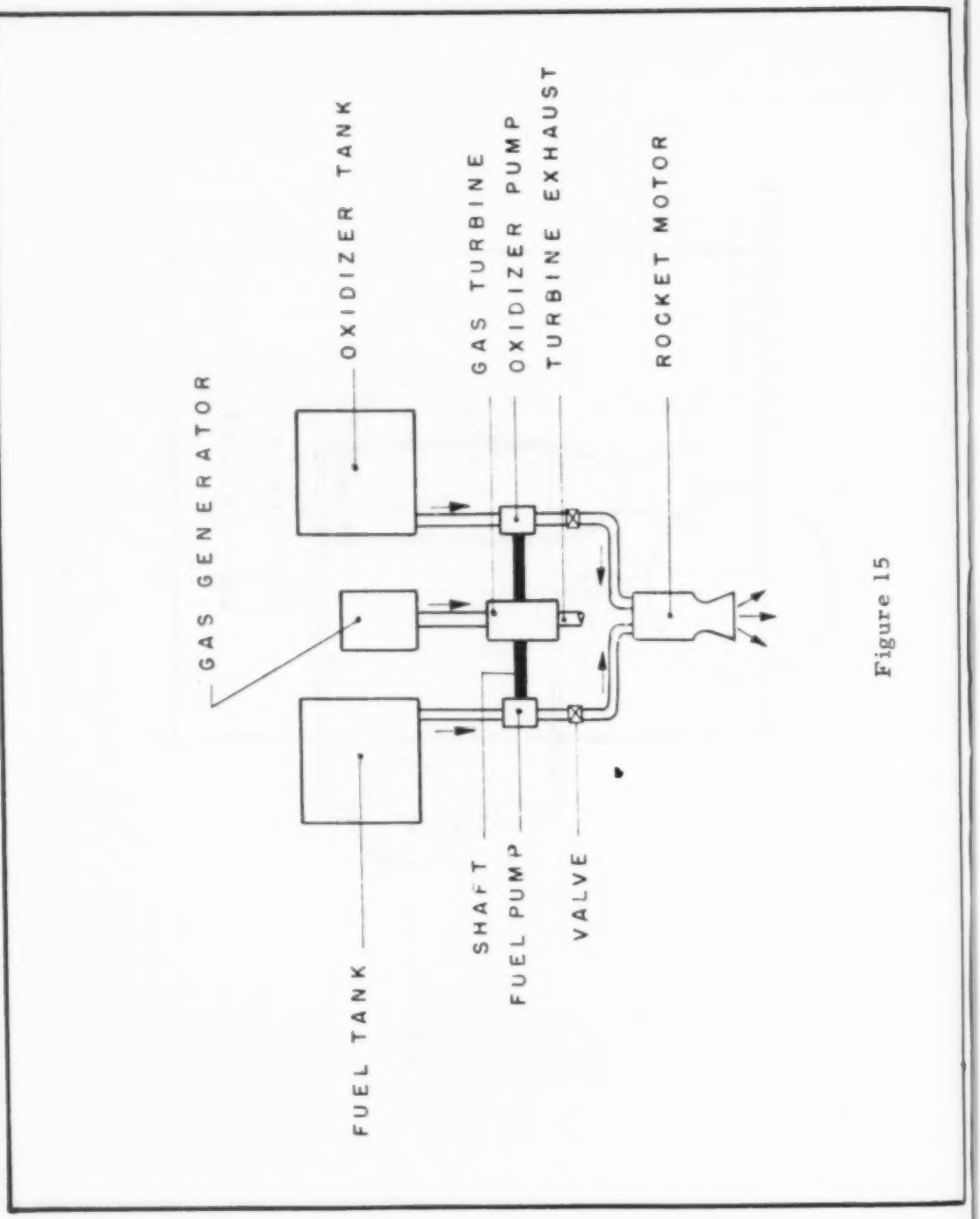

GAS GENERATOR
OXIDIZER TANK
GAS TURBINE
OXIDIZER PUMP
TURBINE EXHAUST
ROCKET MOTOR
FUEL TANK
SHAFT
FUEL PUMP
VALVE

Figure 15

KINEMATICS OF STEERING AN AUTOMOBILE

S. R. UNIECHOWSKI
Associate Professor, Engineering Mechanics
College of Engineering, University of Detroit

This paper represents an attempt to put into equations of kinematics what, to the best of my knowledge, with the exception of the tractrix, has not yet been done in this country. Despite the fact that the main features of the results are generally known to every good automobile or truck driver by intuition or experience, it is hoped that the derivations and results may prove to be enlightning to automotive research personnel and students of mechanical engineering.

After the introduction, dealing with the contact of two solid bodies, the paper is divided into three parts:

I Calculation of the angular velocity of each wheel during turn.
II Kinematic study of the steering mechanism.
III Trajectory of the center of the rear axle when, during a turn, the center of the front axle moves with constant velocity on A:
  (A) straight line
  (B) circular path

Introduction

The contact of two solid bodies.

Let A and B represent two solid bodies moving in such a way that the contact between them is constantly maintained (Figure 1) and let O be one of the points of contact. The motion of a relative to B may be resolved at O into a rotational vector Ω and a couple of moment C. The rotational vector Ω may further be resolved into component vectors Ω_n, normal at O to both surfaces and Ω_t tangent to them. From the above resolution it is evident that the total relative motion A/B results from:

(1) the rotation Ω_t about an axis tangent to both surfaces, which is called rolling at point O;
(2) the rotation Ω_n about a normal, which is defined as pivoting about O;
(3) the translation V, which constitutes the sliding or slipping at point O.

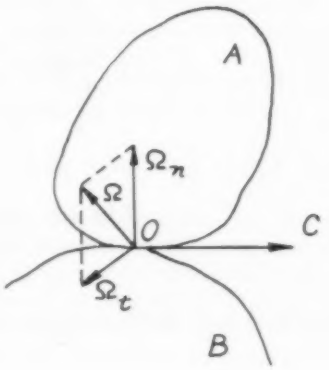


Fig. 1.

When the motion A/A is such that the velocity of translation V is always zero, the bodies roll and pivot one on the other. The motion A/B may be called in this case an

instantaneous rotation. If sliding occurs, the velocity V constitutes the relative velocity with respect to B of the element of contact M at C of body B. It is also evident that the component relative velocities M/A and A/B and their resultant must lie in the tangent plane common to both bodies. The velocity of translation V being their difference must also lie in that plane.

Part I. Calculation of the angular velocities $\omega_1, \omega_2, \omega_3, \omega_4$ of each wheel during a turn.

When there is no sliding between two solids in contact, the point of contact describes curves with the same velocity on each body. These curves must obviously be tangent to each other. In an ideal case (approximated in practice) the slipping at each wheel must be zero (velocity of translation equal zero). If so, the circumferences of all four wheels must be respectively tangent to the trajectories of the four points on contact A_1, A_2, A_3, and A_4.

(1) Let us assume first that during the turn, the steering wheel is immobilized in a certain position. In this case, the points A_1, A_2, A_3 and A_4 are rigidly connected to the chassis. The motion of the chassis being parallel to a horizontal plane, the normals to the trajectories of the points of contact must be concurrent at a certain point C, which is the intersection of the instantaneous axis of rotation Z with the plane of the road. It is evident that the above normals are also normals to the circumferences of all four wheels. This proves that to have simultaneously nonsliding at the point of contact of each wheel with the road, the rear axle and the two front axles must be constantly concurrent in the horizontal projection at a certain point O. Point O is situated on the instantaneous axis of rotation Z, perpendicular to the ground and passing through the point C.

(2) Let us now suppose that the position of the steering wheel is changed during the turning of the car. The points of contact A_3 and A_4 of the front wheels are in this case no longer rigidly connected to the chassis. The relative velocities of these points with respect to the chassis are tangent to the wheels and have the same direction as the resulting velocities with respect to the road. The velocities of translation must be also tangent to the wheels. From the above, we can see that the results obtained in case (1) are also correct in this case.

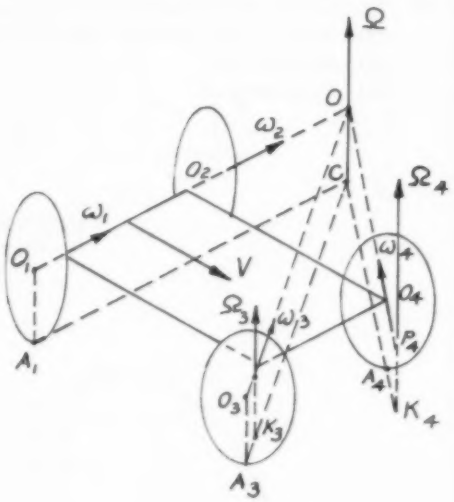


Fig. 2

Let V be the velocity of the center of the rear axle, Ω_3 and Ω_4 the angular velocities of the two front axles with respect to the frame, K_3 and K_4 the horizontal projections of the steering pivots and R the radius of each wheel. If Ω_3 and Ω_4 are known, the angular velocities ω_1, ω_2, ω_3 and ω_4 of all four wheels about their axles as well as the angular velocity of the frame may be obtained in the following manner.

The required condition for each wheel of rolling without sliding will be obtained if the respective point of contact is at that moment the instantaneous center of rotation, i.e., a point of zero velocity. In other words, the relative velocity of the point of contact $\frac{Vai}{Oi}$ must be for the rear wheels equal and opposite to the absolute velocity of their centers. This gives us the first two equations:

$$R\omega_1 = OO_1\,\Omega \ldots\ldots\ldots (1)$$

$$R\omega_2 = OO_2\,\Omega \ldots\ldots\ldots (2)$$

The velocity V of the mid-point of the rear axle is obviously the arithmetic mean between the absolute velocities of O_1 and O_2; therefore,

$$V = \left(\frac{OO_1 - OO_2}{2}\right)\Omega \ldots\ldots\ldots (3)$$

For the front wheels, the velocity of the center of each wheel is the resultant of the velocity due to rotation around the pivot and the velocity due to the rotation of the whole frame; therefore,

$$R\omega_3 = OO_3\,\Omega \pm R\,\Omega_3 \ldots\ldots\ldots (4)$$

$$R\omega_4 = OO_4\,\Omega \mp R\,\Omega_4 \ldots\ldots\ldots (5)$$

where r is the length of each front axle, i.e., the distance from the center of the wheel from the pivot. The signs in equations (4) and (5) depend on the direction in which the steering wheel is being turned.

From those five equations, the angular velocities of each wheel may be calculated. They are different for each wheel and equal only when Ω_1, Ω_3 and Ω_4 are all zero. When the steering wheel is maintained in a certain position (Ω_3 and Ω_4 are equal to zero), point C is fixed with respect to the chassis and also with respect to the road. The motion of the automobile consists of a rotation about a fixed axis.

In any case, the middle point of the rear axle moves along a certain curve tangent to the longitudinal axis of symmetry of the car. It may be easily seen that the turning radius of the rear of the automobile is smaller than that of the front, the difference increasing with the length of the car.

Part II. Kinematic study of the steering mechanism.

To approximate as far as possible the ideal condition described in part I, the steering mechanism consists of steering arms SP_3 and TP_4 joined by a tie-rod ST. In the initial position, a trapezoid P_3STP_4 is formed (Figure 3). For the sake of simplicity, the foregoing study will be limited to small angles made by the forward axles with their initial position. The following notations will be used:

$$P_3P_4 = B;\quad P_3S = P_4T = D;\quad ST = F;\quad \angle P_4P_3S = \angle P_3P_4T = D$$

It follows that B = F+2d.cosd. Let the distance between the front and rear axles = L. The turning of the steering wheel produces the deformation of the three-link mechanism P_3STP_4 into $P_3S_1T_1P_4$ which in turn produces the rotation of the two front axles through angles θ and ξ, as shown on Figure 3. The tie-rod making an angle β with its initial position. In accordance with part I point C, the intersection of the two front axles in their new position must be located on the line of the rear axle, at a distance L from P_3P_4.

To obtain a relation between the variable angles θ and ξ, let us express first the distance L in terms of θ and ξ. From triangle CP_3P_4:

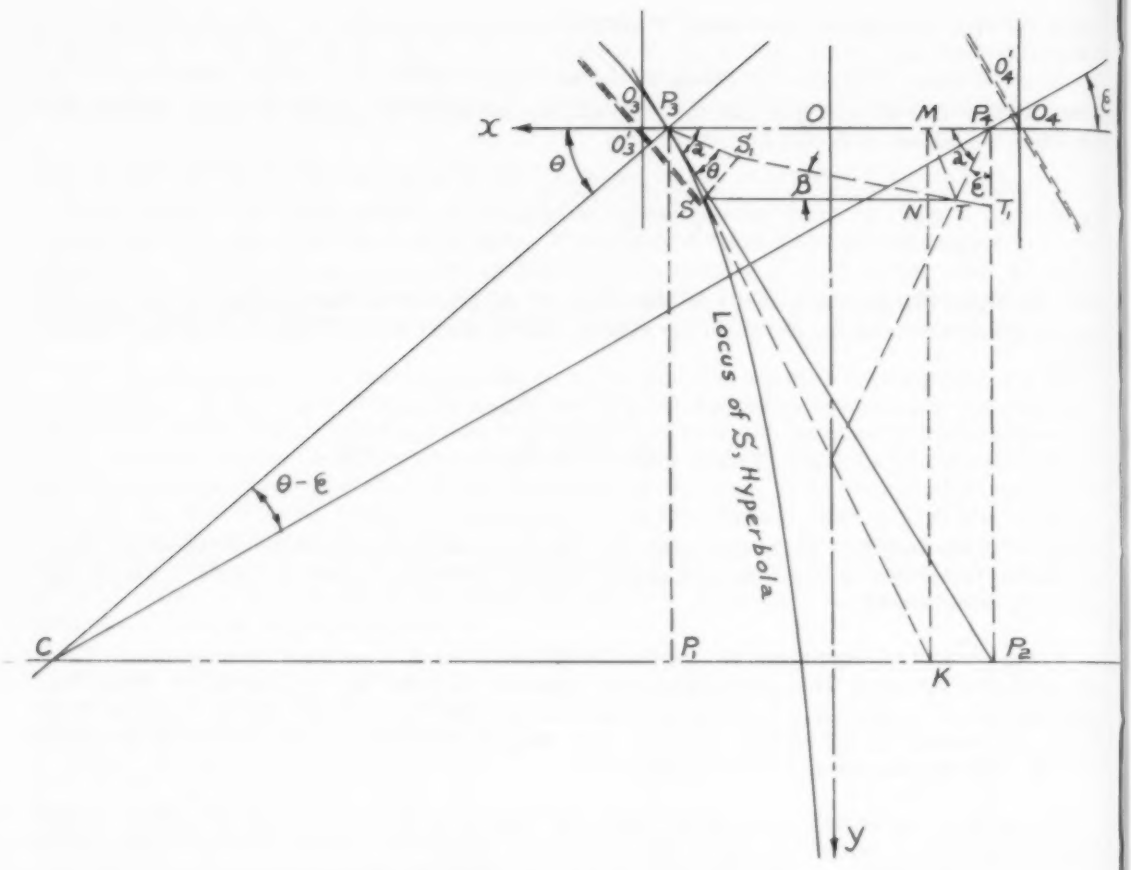


Fig. 3

$$\frac{B}{\sin(\theta-\phi)} = \frac{CP_3}{\sin\phi} \quad ; \quad \text{BUT}$$

$$CP_3 = \frac{L}{\sin\theta} \quad \text{THEREFORE}$$

$$\frac{B}{\sin(\theta-\phi)} = \frac{L}{\sin\theta\,\sin\phi}$$

Giving

$$L = \frac{B \cdot \sin\theta \cdot \sin\phi}{\sin(\theta-\phi)}$$

which for small values of θ and ϕ becomes

$$L = \frac{B \cdot \theta \cdot \phi}{\theta-\phi}$$

To obtain a relation between the variable angles $\beta \cdot \theta$ and ϕ, let us project the three-link mechanism P_3STP_4 into horizontal and vertical directions.

$$F\cdot\cos\beta - B + d\cos(\alpha-\theta) = -d\cos(\alpha+\epsilon) \quad \ldots\ldots(1)$$

$$F\cdot\sin\beta + d\sin(\alpha-\theta) = d\sin(\alpha+\epsilon) \quad \ldots\ldots(2)$$

By squaring and adding equations (1) and (2) ϵ can be eliminated giving an implicit function of the form $F(\beta,\theta) = 0$, which after simplification will yield:

$$F(\beta,\theta) = f^2 + b^2 - 2fb\cos\beta + 2d(f\cos\beta - b)\cos(\alpha-\theta) + 2df\sin\beta\sin(\alpha-\theta) = 0 \quad \ldots\ldots(3)$$

To obtain an explicit relation $\theta = f(\beta)$ the function θ is expanded into a series by Maclaurin's theorem.

From (3) we have

$$\theta' = \frac{d\theta}{d\beta} = -\frac{\partial F/\partial\beta}{\partial F/\partial\theta}$$

$$= -\frac{2fb\sin\beta - 2d(f\cos\beta - b)\cos\alpha - 2fd\sin\beta\sin(\alpha-\theta)}{2d(f\cos\beta-b)\sin(\alpha-\theta) - 2fd\sin\beta\cos(\alpha-\theta)}$$

Giving for $\beta = 0$, $\theta'_{\beta=0} = -\dfrac{fd\sin(\alpha-\theta)}{d(f-b)\sin(\alpha-\theta)}$

or $\theta'_{\beta=0} = \dfrac{-f}{f-b} = \dfrac{f}{2d\cos\alpha}$

Using a similar process for ϵ, it may easily be verified that

$$\epsilon'_0 = \theta'_0 = \frac{f}{2d\cos\alpha}$$

To obtain

$$\theta'' = \frac{d^2\theta}{d\beta^2}$$

The following formula for an implicit function is used:

$$\frac{d^2\theta}{d\beta^2} = -\frac{\left(\dfrac{\partial F}{\partial\theta}\right)^2\dfrac{\partial^2 F}{\partial\beta^2} - 2\dfrac{\partial F}{\partial\beta}\cdot\dfrac{\partial F}{\partial\theta}\cdot\dfrac{\partial^2 F}{\partial\beta\cdot\partial\theta} + \left(\dfrac{\partial F}{\partial\beta}\right)^2\dfrac{\partial^2 F}{\partial\theta^2}}{\left(\dfrac{\partial F}{\partial\theta}\right)^3}$$

Taking into account that when

$$\beta = 0;\ \theta = 0$$

Also, we finally obtain

$$\theta''_0 = \frac{d^2\theta}{d\beta^2} = \frac{fb}{2d^2\sin 2\alpha}$$

and again by a similar process for ϵ

$$\epsilon_o'' = -\theta_o'' = -\frac{f b}{2 d^2 \sin 2\alpha}$$

The expansions of

$$\theta = f(\beta)$$

and of

$$\epsilon = f_1(\beta)$$

will therefore give

$$\theta = \frac{f}{2d\cos\alpha}\beta + \frac{f b}{4 d^2 \sin 2\alpha}\beta^2 + \cdots\cdots$$

and

$$\epsilon = \frac{f}{2d\cos\alpha}\beta - \frac{f b}{4 d^2 \sin 2\alpha}\beta^2 + \cdots\cdots$$

It should be noted that when

$$\beta = 0; \quad \theta = \epsilon = 0$$

and therefore the first terms in both expansions for θ and ϵ equal zero.
For small values of the angles θ and ϵ the equation (A),

$$L = \frac{b \cdot \theta \cdot \epsilon}{\theta - \epsilon}$$

tends to:

$$b\left[\frac{f}{2d\cos\alpha}\beta + \frac{f b}{4 d^2 \sin 2\alpha}\beta^2\right]\left[\frac{f}{2d\cos\alpha}\beta - \frac{f b}{4 d^2 \sin 2\alpha}\beta^2\right]$$

which, after simplification and omitting terms of negligible magnitude, gives finally

$$L = f \cdot \tan\alpha$$

On the other hand it is clear from Figure 3 that the above result constitutes the essential requirement for the ideal condition described in part I. The line of the rear axle must therefore be an asymptote to the locus of point C (the intersection of the lines of the two front axles). To obtain in practice the relation $L = f \cdot \tan\alpha$, the following condition is necessary:

In the parallelogram P_3 STM, obtained by drawing a line TM parallel to SP_3, the line P_3 S produced to the intersection with the rear axle must intersect it at a point K, located on the perpendicular dropped from M on the line of the rear axle.

There is no limit to the number of possible solutions. The locus of point S may be obtained as follows:

For coordinate axes as shown on Figure 3, from similar triangles P_3 MK and SNK, observing that $P_3 M = ST = f$

$$\frac{SN}{P_3 M} = \frac{NK}{MK} \quad \text{or} \quad \frac{x - (f - b/2)}{f} = \frac{L - y}{L}$$

But $X = \frac{f}{2}$ giving

$$\frac{X + (2X - b/2)}{2X} = \frac{L - y}{L}$$

and

$$3LX - \frac{b}{2} = 2X\left(\frac{L-y}{L}\right)$$

yielding further

$$LX + 2Xy = \frac{bL}{2}$$

and finally

$$X\left(y + \frac{L}{2}\right) = \frac{bL}{4}$$

This is a hyperbola for which the Y-axis (longitudinal axis of the automobile) is an asymptote. The diagonal $P_3 P_2$ of the rectangular $P_3 P_4 P_2 P_1$ is tangent to this hyperbola at the point P_3.

In practice the point S is chosen comparatively near to P_3 (small values of D) to allow for the values of angles θ and φ up to about 40^0. Taking for instance $\theta = 40^0$ and $\varphi = 35^0$, it may be seen that the error in formula (A), due to the approximation for small angles (angles taken instead of their sines) is only about 7 per cent.

Part III. Trajectory of the center of the rear axle when, during a turn, the center of the front axle moves with constant velocity on A:

(A) Straight line.

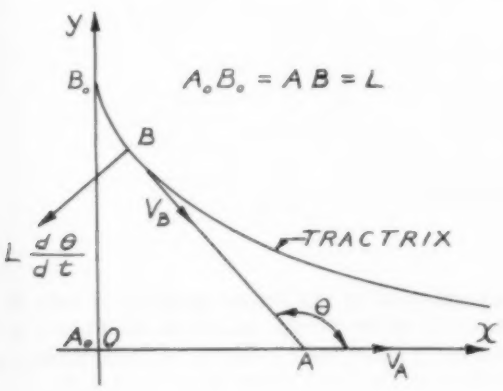


Fig. 4

Let A_0 and B_0 represent respectively the centers of the front and rear axles in their initial positions and let L be the distance A_0B_0. Point A moves to the right along the x-axis, while point B follows a certain path, which has to be determined. After a certain time, the centers of both axles will be in the position AB, as shown on Figure 4. Let V_A and V_B represent respectively the velocities of points A and B and let θ be the angle formed at a given moment between AB and the x-axis. V_A must be directed along O_A and V_B along AB. From elementary relations of kinematics

$$V_B = V_{B/A} \longrightarrow V_A \quad \ldots\ldots (1)$$

Projecting the vectors in the direction of V_B, perpendicular to V_B and observing that

$$V_{B/A} = L\frac{d\theta}{dt}$$

It is immediately obtained: $L \frac{d\theta}{dt} - V_A \sin\theta = 0$ ..... (2)

giving $V_A = \frac{L}{\sin\theta} \frac{d\theta}{dt}$ ..... (3)

Now let us project the equation (1) in the directions of the coordinate axes:

$$(V_B)_x = \frac{dX}{dt} = V_A - L \frac{d\theta}{dt} \sin\theta \quad \ldots\ldots (4)$$

and

$$(V_B)_y = \frac{dY}{dt} = -L \frac{d\theta}{dt} \cos\theta \quad \ldots\ldots (5)$$

Replacing V_A in (4) by its value from (3) and simplifying, we obtain the following set of equations:

$$dX = L\left(\frac{d\theta}{\sin\theta} - \sin\theta\, d\theta\right) \quad \ldots\ldots (6)$$

$$dY = -L \cos\theta\, d\theta \quad \ldots\ldots (7)$$

By chosing for origin O the point A_0 (t=O) the integration of the last set of equations yields:

$$X = L\left[\log_e\left(\tan\frac{\theta}{2}\right) + \cos\theta\right] \quad \ldots\ldots (8)$$

$$Y = -L \sin\theta \quad \ldots\ldots (9)$$

which represent a parametric set of equations of a tractrix in its simplest form, most convenient for its construction. Equations (8) and (9) may easily be converted into the usual form, namely:

$$X = t - L \tanh\frac{t}{L}$$

$$Y = L \operatorname{sech}\frac{t}{L}$$

where t is the x- intercept of the tangent to the tractrix.

(B) Circle of radius R.

As in (A), let A represent the center of the front axle and B the center of the rear axle of the automobile. For the origin selected in the center O of the circle, as shown on Figure 5, ρ and θ are the polar coordinates of point B, the trajectory of which has to be investigated. From triangle OAB, OA = R and OB = ρ; AB = L

$$R^2 = L^2 + \rho^2 - 2L\rho\cos\alpha \quad \text{or} \quad \cos\alpha = \frac{L^2 + \rho^2 - R^2}{2L\rho} \quad \ldots\ldots (1)$$

where α represents the angle made by a tangent to the trajectory of point B with the radius vector ρ. Therefore

$$\tan\alpha = \frac{\rho\, d\theta}{d\rho} \quad \ldots\ldots (2)$$

The angle α can be eliminated from (1) and (2) by taking

$$\tan\alpha = \frac{\sqrt{1-\cos^2\alpha}}{\cos\alpha}$$

or

$$\frac{\rho\, d\theta}{d\rho} = \frac{1-\sqrt{\dfrac{L^2+\rho^2-R^2}{2L\rho}}}{\dfrac{L^2+\rho^2-R^2}{2L\rho}} = \frac{\sqrt{4L^2\rho^2-(L^2+\rho^2-R^2)^2}}{L^2+\rho^2-R^2}$$

This leads to a differential equation:

$$d\theta = \frac{\sqrt{4L^2\rho^2-(L^2+\rho^2-R^2)^2}\, d\rho}{(L^2+\rho^2-R^2)\,\rho} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

$$d\theta = \frac{\sqrt{\left[R^2-(L-\rho)^2\right]\cdot\left[(L+\rho)^2-R^2\right]}\, d\rho}{(L^2+\rho^2-R^2)\,\rho} \quad \ldots\ldots\ldots\ldots (4)$$

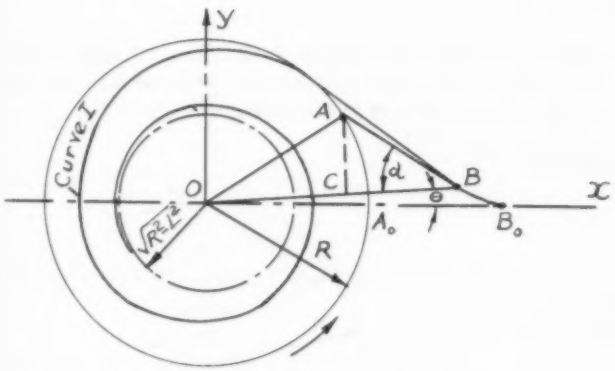


Fig. 5

The last equation may be integrated by means of a number of substitutions leading finally to the equation of the trajectory of point B in terms of elementary transcendental functions. The above procedure involves a considerable amount of labor and it would seem more practical to discuss directly the equation (3) or (4) and the change of sign of $\frac{d\theta}{d\rho}$. It is evident that $\frac{d\theta}{d\rho}$ will be equal to zero in the initial position or some extreme positions when the points O, A, and B will be on a straight line. For R > L, $\frac{d\theta}{d\rho}$ becomes infinite for

$$\rho = \sqrt{R^2-L^2}$$

(I) R > L. Let us assume an initial position $\rho_o \cdot \theta_o$ for $\rho_o > \sqrt{R^2-L^2}$

and let A turn in the positive (counterclockwise) direction. The derivative $\frac{d\theta}{d\rho}$ is negative at the initial position and continues to be negative until its value becomes zero or infinite.

Therefore, θ increases while ρ decreases and

$$\theta-\theta_o = -\int_{\rho_o}^{\rho} \frac{\sqrt{\left[R^2-(L-\rho)^2\right]\cdot\left[(L+\rho)^2-R^2\right]}\, d\theta}{(L^2+\rho^2-R^2)\,\rho} \quad \ldots\ldots\ldots\ldots (5)$$

giving for θ a finite value until ρ decreasing in magnitude has attained the critical value

$$\sqrt{R^2-L^2},$$

for which ρ becomes infinite.

It may be proved that for

$$\rho=\sqrt{R^2-L^2},$$

the primitive function becomes also infinite. This proof is given in appendix I. Therefore the decreasing values of ρ a circle of radius

$$\sqrt{R^2-L^2}$$

becomes a curve asymptotic to our trajectory. For the increasing values of ρ (point A turns in the negative, i.e., clockwise direction). θ decreases until the extreme case $\frac{d\theta}{d\rho} = 0$, when the curve presents a cusp with a tangent passing through the origin at O. The trajectory for case (I) is shown on figure 5 (curve I). In practice, this case represents what occurs during a left hand turn encircling a roundabout at the intersection of two highways (Figure 6).

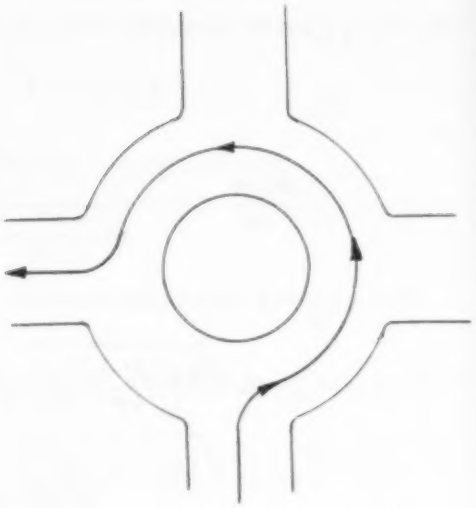


Fig. 6

The next interesting case, corresponding in practice to a U-turn, occurs when

$$\rho<\sqrt{R^2-L^2}$$

and the initial position ρ=R - L. The general analogous to case (I). The curve is bounded by a circle of radius

$$\sqrt{R^2-L^2}$$

which is asymptotic to the trajectory. The curve presents a cusp on a circle of radius R - L.

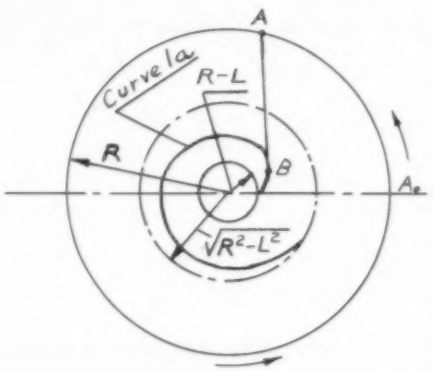


Fig. 7

(II) R < L. In this case, which does not occur in practice and is therefore purely theoretical, the integrand does not become infinite because ρ can vary only between L + R and L - R. The trajectory presents two cusps at distances L + R and another at L - R from the center O of a given circle (curve II, Figure 8).

(III) R = L. In this case:

$$\frac{d\theta}{d\rho}=\frac{\sqrt{4L^2-\rho^2}}{\rho^2}$$

becomes infinite for $\rho = 0$ The primitive function:

$$\theta - \theta_0 = -\left[\frac{\sqrt{4L^2 - \rho^2}}{\rho} - \arcsin\frac{\rho}{2L}\right]_{\rho_0}^{\rho}$$

also becomes infinite for $\rho = 0$. Therefore the trajectory of point B curls around O, which is an asymptotic point of the trajectory.

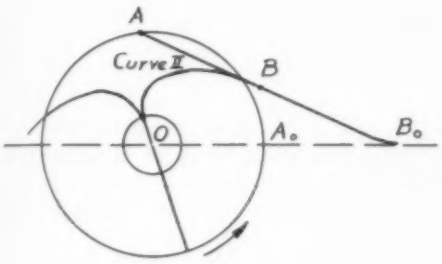


Fig. 8

All three cases can be demonstrated on a model, the results obtained being in full agreement with the calculations.

Appendix I

To prove that in equation (5), part III, section (b), the primitive function $F(\rho)$ also becomes infinite for

$$\rho = \sqrt{R^2 - L^2}$$

when the integrand is infinite, let us consider another function $f_1(\rho)$ which is dominant with respect to the integrand $f(\rho)$.

Selecting

$$f_1(\rho) = \frac{d\rho}{\rho - \sqrt{R^2 - L^2}}$$

which is of the same order of magnitude for

$$\rho = \sqrt{R^2 - L^2}$$

as $f(\rho)$, we may say that if the primitive function of $f_1(\rho)$, $F_1(\rho)$ becomes infinite for the above value of ρ, $F(\rho)$ must also be infinite.

$$F_1(\rho) = \int_{\rho_0}^{\rho} \frac{d\rho}{\rho - \sqrt{R^2 - L^2}} = \left[\log_e\left(\rho - \sqrt{R^2 - L^2}\right)\right]_{\rho_0}^{\rho}$$

becomes infinite for

$$\rho = \sqrt{R^2 - L^2}$$

therefore, $F(\rho)$ must also be infinite.

The above results may be verified directly by obtaining $F(\rho)$ by integration of equation (4) or (3) from part III, section b. Since the procedure is long and laborious, only the main steps will be indicated.

$$d\theta = \frac{\sqrt{4L^2\rho^2 - (L^2+\rho^2-R^2)^2}\, d\rho}{(L^2+\rho^2-R^2)\rho}$$

where L and R are constants. Use the substitution

$$L^2+\rho^2-R^2 = 2L\rho\cos\alpha$$

then

$$d\rho = -\frac{L\rho\sin\alpha\, d\alpha}{(\rho - L\cos\alpha)}$$

in which

$$\alpha = \arccos\frac{L^2+\rho^2-R^2}{2L\rho}$$

Hence

$$d\theta = -\frac{L\sin^2\alpha\, d\alpha}{(\rho - L\cos\alpha)\cos\alpha}$$

From triangle OAC in Figure 5

$$OC = \rho - L\cos\alpha = \sqrt{R^2 - L^2\sin^2\alpha}$$

giving

$$d\theta = \frac{L\sin^2\alpha\, d\alpha}{\cos\alpha\sqrt{R^2 - L^2\sin^2\alpha}}$$

The substitution $\frac{L}{R}\sin d = \sin u$ yields

$$d\theta = \frac{-R^2\sin^2 u\, du}{L^2 - R^2\sin u}$$

Finally, substituting u = sin V and v = arc sin u, we obtain, i.e.,

$$d\theta = -\frac{R^2V^2\cos V\, dV}{L^2 - R^2V}$$

The last equation may be expanded into a series:

$$d\theta = -\frac{R^2V^2\left(1 - \frac{V^2}{2!} + \frac{V^4}{4!} - \frac{V^6}{6!} + \cdots\right)dV}{(L+RV)(L-RV)}$$

and integrated by partial fractions, giving a certain number of transcendental functions which become infinite for

$$\rho = \sqrt{R^2 - L^2}$$

I should like to acknowledge the assistance given me in the preparation of this paper by Professors L. S. Johnston and J. Lubelfeld of the Department of Mathematics of the University of Detroit.

PROBLEMS IN SHOCK & VIBRATION CONTROL

R. N. JANEWAY
Head - Dynamics Research Department
Engineering Division
Chrysler Corporation

INTRODUCTION

The primary purpose of this paper is to review the underlying dynamic principles that apply to all types of suspension and vibration isolation problems as well as to the suspension of vehicles. However, specific attention has been given to the practical application of these principles to the design of automobile and other vehicle suspensions.

Special emphasis has been placed on the problem of suspension damping, which, in the final analysis usually determines the overall operating performance for a given static deflection. Basically, this situation stems from the fact that damping, although indispensable under certain critical conditions, is quite undesirable under other, and usually more prevalent conditions, especially in vehicles. Consequently, the optimum damping characteristic for any given suspension necessarily represents a compromise whereby the minimum penalty is paid for the essential control.

In considering the available types of damping, the characteristics of coulomb or constant friction damping have been extensively explored, and compared with those of viscous damping. Although dry friction damping devices are the oldest known type, they have proved, in the past, to be generally unsatisfactory, both as to performance and life. Their use of friction materials with high static breakway and low dynamic friction coefficients clearly made it impossible to approach the constant friction ideal of coulomb damping. Furthermore, these devices suffered from a notable lack of precision in the relationship of their component parts which made close calibration of the friction impossible. In the meantime, the expanding need for suitable damping devices has been largely filled by various types of hydraulic mechanisms.

Aside from the practical deficiencies of previous friction devices, dry friction damping has also suffered because the fundamental characteristics of coulomb damping have not been generally understood. Most text books on vibration are content with a superficial treatment limited to the condition of free vibration. On the other hand, viscous damping where the damping force varies in proportion to velocity, has been elaborately treated. Consequently, the fundamental characteristics of viscous damping are very well established. One of the contributing factors for this neglect of coulomb damping is the fact that it is exceedingly disagreeable to handle mathematically because it cannot be expressed as a continuous mathematical function. In contrast, viscous damping can be expressed as a continuous function in a linear differential equation and thus involves no frustrations to the mathematician.

Two comparatively recent developments have stimulated progress in the field of coulomb damping. First, a friction device is now available which embodies in practical form the ideal constant friction or coulomb damping. Consequently, a thorough knowledge of the characteristics of such damping has now become an objective of much more than academic interest. Second, the availability of electronic calculating machines has made it possible to solve quickly the complex mathematical equations involved, that previously defied long hand methods. Although the results to date are far from exhaustive, it is hoped that their presentation here will be useful in a wide variety of problems involving the control of shock and vibration.

The invaluable assistance of Mr. F. F. Timpner, project engineer of the Chrysler Dynamics Research Department, both in carrying out the calculations and in the presentation of this paper, is gratefully acknowledged.

I. UNDERLYING PRINCIPLES

Basic Importance of Static Spring Deflection

Among the underlying principles, there is one that cannot be stressed too strongly, namely, that of static deflection of the elastic elements of a suspension. First of all, the greater the static deflection the lower is the natural frequency of oscillation and it is fundamental that the lower the natural frequency relative to the frequency of forced vibration the smaller will be the transmission of the vibration. Also, it will be shown that the frequency is a potent factor in determining human tolerance of vibration.

It is also quite evident that the dynamic force imposed on a sprung mass by a given spring deflection, X, from the static position is going to vary directly as the deflection rate of the spring, or $F = KX$. The resulting acceleration of the sprung mass will be $a = \frac{F}{M} = g\frac{KX.}{W}$ Therefore, the ratio of the dynamic spring deflection, X, to the static deflection, W/k, defines the acceleration directly in gravity units. Stated another way, the acceleration imposed by a given dynamic spring deflection clearly will vary inversely as the static deflection.

In dealing with the effects of vibration on structures or inanimate vehicle lading it is logical to use force transmission to evaluate the result. Likewise, we rightly measure shock in terms of the acceleration transmitted which, of course, is a direct measure of force. However, when we get into the question of passenger comfort in a vehicle, we find that acceleration is no longer sufficient measurement for purposes of evaluating human reaction. Careful study of the results of investigations into the subjective responses of human beings to vibration shows that "jerk" or the rate of change of acceleration comes nearer to being a criterion of comfort or discomfort. (1) This holds only in the frequency range of 1 to 6 cycles per second, and above that we find that other criteria take over. However, the 1 to 6 cycle range is precisely where the principle disturbances occur in vehicles and, therefore, it is highly pertinent to examine the implications of this comfort criterion.

If we consider a sine wave vibration we find that the maximum jerk is a function of the product of amplitude, a, and the cube of frequency, f; or $j_m = 8\pi^3 a f^3$. Since the frequency varies inversely as the square root of static deflection, δ, or $f = \frac{1}{2\pi}\left(\frac{g}{\delta}\right)^{\frac{1}{2}}$, the maximum jerk must be related to the static deflection as $j_m = \frac{g^{1.5}\alpha}{\delta^{1.5}}$.

Therefore, if we derive the limiting amplitude required to maintain a constant maximum jerk value, this becomes $a_m = \frac{j_m \delta^{1.5}}{g^{1.5}}$, or the limiting amplitude is proportional to the 1 1/2 power of static deflection. This relationship is shown in the first chart, Fig. 1. What we have here is both the comfort limited amplitude based on constant jerk and the oscillation frequency, both plotted against the effective statis spring deflection. When we say effective deflection we mean, not the deflection of the spring from its free condition to the loaded condition, but the deflection obtained by dividing the load by the actual spring rate under the loaded condition.

If we take 2" as being the maximum vertical amplitude to be accommodated in an automobile, the curve shows that to maintain a tolerable level of disturbance without discomfort, for short duration, we must provide a static spring deflection of approximately 9". This means a bouncing frequency of about 60 cycles per minute, as is shown by the frequency curve

in Fig. 1. This is precisely the frequency level that we are now approaching in current automobile suspensions.

It is evident, then, that as we increase the static spring deflection we can tolerate, because of lower frequency, a rapidly increasing amplitude. Looking back on the period, less than twenty years ago, when static deflections of 3" to 4" represented the maximum available in automobiles, it becomes plain that the passengers then were being subjected to vibrations which extended far into the discomfort zone. As Chart 1 shows, the comfort limited amplitude at a static deflection of 4" is little more than 1/2". Therefore, even at a bouncing amplitude of 2" the tolerable limit would be exceeded by a factor of 4 to 1.

Relation of Wheelbase to Mass Distribution

The next factor I would like to discuss is that of the sprung mass distribution, where the mass is supported on two spaced suspension elements. Just considering the vertical longitudinal plane, such a body has, of course, two degrees of freedom, namely, angular motion about its center of gravity and vertical motion of the center of gravity. Now the angular motion which is perhaps even more important practically than the vertical motion, is very largely determined by what we call the dynamic index, which is the ratio $\frac{i^2}{AB}$, where i = the radius of gyration and A & B are the distances from the center of gravity to the wheel centers. In most automobiles as they are today, that index falls between .75 and 1. Twenty years ago, practically all automobiles had index values not exceeding .60 to .70. However, since the pioneering design of the Chrysler Airflow in 1934, there has been a gradual change throughout the industry toward greater overhang which has increased the radius of gyration relative to the wheelbase. But although automobiles today are in a much higher range of dynamic index, most cars are still far from having an index of 1, which we consider the ideal to be approached.

To bring out the important quantitative effect of this relationship, Appendix 1 gives a mathematical analysis of the dynamics of the angular acceleration of the sprung mass. Two possible interpretations can be applied to this derivation. One is that the pivot is rigidly supported, in which case we solve for the reaction force developed at the pivot. The alternative assumption is that the pivot is flexibly supported and, therefore, that the effective instant center of rotation is not necessarily located at the point of support. It will be noted that the result is obtained by the solution of two simultaneous equations, one expressing the angular acceleration about the pivot, and the other, expressing the angular acceleration about the center of gravity. Since the angular acceleration must be the same about both points, in a rigid body, we get a simple solution in terms of the radius of gyration and the locations of the points of support. Thus,

if R = the reaction force at one point of support (assumed rigid)
F = applied force at the other point of support
A = distance of C.G. from point of force application
B = distance of C.G. from pivot point
i = radius of gyration of mass

Then, $\frac{R}{F} = \frac{AB - i^2}{i^2 + B^2}$

Now, if the reaction force, R, is to be zero, it is obvious by inspection that the product AB must be equal to i^2. That is the only way in which the reaction can be eliminated, and corresponds to the condition for which the dynamic index is 1.

The relationship becomes even simpler if we take the special case where the center of gravity is equidistant from the supports. Then A = B and the equation for the reaction force ratio, $\frac{R}{F}$ reduces to $R/F = \frac{1 - p}{1 + p}$ where p = dynamic index = i^2 / A^2. Here again, when p = 1, the reaction force becomes zero.

In the actual vehicle suspension, if we consider that the force is applied at one spring and that the pivot is located at the other spring center, obviously we have a pivot that is flexibly supported. Consequently, unless the dynamic index is equal to 1, angular acceleration cannot take place about the spring as a pivot. For, in order for a reaction force to exist at the pivot, deflection of the spring must first take place. Therefore, the instantaneous center of rotation must lie at some point other than the spring support. It is easily seen from the equations where that point would be, namely, at a distance from the center of gravity which is equal to $\frac{i^2}{L - R}$. This point is also readily identified as the center of percussion of the mass with respect to the point of force application.

In the curves of Chart 2, quantitative relationships are given for the two conditions; first, in the case of a rigid pivot, the relative force corresponding to any value of dynamic index, and second, in the case of a flexible support, where the instantaneous center of rotation will be located. For example, we see that for a dynamic index of .70, the reaction force at a rigid support, given by the lower curve, is 17-1/2% of the applied force. On the other hand, if we consider the actual condition of a flexible support, the upper curve shows that the instantaneous center of rotation would be displaced from the spring center, a distance equal to 12% of the spring spacing. In other words, if we have a wheelbase of 120 inches and apply a force at the front spring, the center of rotation will be displaced 14 inches forward of the rear wheel center. Of course, this means that the rear spring must be deflected as a result of reaction from every disturbance of the front spring. In turn, when a force is applied at the rear spring, rotation will take place about a point 14" to the rear of the front spring, in the example being considered. It is quite evident that this action is undesirable for two reasons. First, it tends to excite sustained pitching oscillation about a center between the springs; and second, it decidedly augments the deflection of the rear springs as the car traverses a road irregularity.

Suppose the front wheels hit a road bump, compressing the front springs and imposing an upward angular acceleration on the sprung mass. If the instantaneous center of rotation is forward of the rear spring, clearly the motion of the sprung mass at the rear spring is such as to compress the spring. This action takes place before the rear wheels strike the bump. When the direct displacement of the rear wheels subsequently takes place, the resultant spring compression is necessarily greater than it would have been if no reaction had resulted from the front spring disturbance. This succession of events, aggravated by relatively stiff springs, led inevitably to the rear seat passengers being tossed clear of the seat in the "pre-Airflow" era of automobile design. Even with the present refinement in suspension, the disagreeable effects of improper weight distribution on automobile ride can still be felt as well as measured.

The Fallacy of Long Wheelbase

Now, what does this mean in terms of automobile design? In general, the practical realization of a high dynamic index calls for a determined effort to meet two objectives; first, to locate concentrated masses as far outboard as possible, and so realize the maximum radius of gyration, and second, to restrict the wheelbase in keeping with the radius of gyration. As a rule, the larger and heavier the car the greater the tendency of the wheelbase to outrun the radius of gyration, although this can be avoided by strict adherence to the objective of high dynamic index. Therefore, it is too often the case that the inherent riding quality deteriorates as the wheelbase and cost of the car increase. Of course, this tendency can be partially offset, so far as the passenger is concerned, by softer springs, better cushions, etc. The fact remains that contrary to common belief, long wheelbase in itself is a liability, not an asset in achieving optimum riding quality.

One other fundamental factor in connection with the dynamic index should be mentioned. This is the effect of the center of gravity location relative to the spring centers. It is readily seen that the spacing, A + B, between centers will increase continuously for a constant value

of the product AB, as the center of gravity is shifted away from the central position. Conversely, for a fixed spacing, A + B, the product, AB, becomes smaller the more unequal A and B become. It follows that unequal weight distribution between front and rear wheels tends to improve the dynamic index, for a given wheelbase and radius of gyration. The gain from this source is necessarily limited by the requirements of traction in acceleration, braking and handling. Still, the difference between equal weight distribution and, say, a 58 to 42 ratio of front to rear sprung weight, amounts to about 3% change in the index in the right direction.

Thus, we see the unmistakeable trends in automobile design toward long body overhang front and rear, powerplant located as far forward as possible, and an increasing percentage of total weight on the front wheels, all of which are necessary in maintaining a high dynamic index.

II. DAMPING - GENERAL CONSIDERATIONS

Excitation Dictates Control Requirements

We come now to the subject of damping, which of course is one of the most vital to the practical functioning of any suspension system. For, even if the optimum conditions have been met with regard to spring deflections and weight distribution, it is essential to provide the type and degree of control that will give the best overall result under actual operating conditions.

In general, the damping requirements are determined by the nature of the excitation to which the system will be subjected; and, in particular, to the possibilities of the excitation inducing resonant oscillation of the system. Let us consider various types of excitation to see how they influence the control requirements. The first and simplest condition is steady state excitation over a limited frequency range, or possibly at a constant frequency. In a case of this kind, no damping is really needed as long as an isolating mounting can be designed to have a natural frequency which is far enough away from the operating frequency so that resonance is never approached. Moreover, better isolation is obtained by virtue of the absence of damping. However, if steady state excitation can be imposed over a wide, and possibly uncontrollable range of frequency, then the damping requirement becomes most severe. An example of steady state resonance is the vibration of unsprung masses of a car at the critical speed on a washboard road. This is one of the few cases in a road vehicle where it is possible to approach a steady state resonant excitation and, of course, it points up the need for adequate damping of the unsprung masses.

Another important case to consider is excitation by regularly repeated transient impulses of short duration occurring relative to the period between impulses. Now here it is possible to get resonance due to the frequency of the impulses. However, between impulses the vibration is no longer forced but free so all that is needed is enough damping to bring the mass to rest before the next impulse occurs. A good example is a railroad vehicle traversing regularly spaced rail joints, each of which produces a transient impulse of relatively short duration. It is evident that there will be a critical speed at which the frequency of the impulses is identical with the natural frequency of oscillation. To take a specific case, if the sprung mass has a natural frequency of 2 cycles per second and the rail length is 39 ft., the critical speed will occur at 78 ft. per second, or 53 miles per hour. If the duration of the forced wheel deflection corresponds to a travel of 6 ft., the sprung mass will have $\frac{33}{39} = 85\%$ of one cycle of free vibration in which its motion can be completely damped.

The speed of 53 mph., cited above, is about the lowest critical speed likely in freight equipment. In passenger equipment, it is usually about one-half as great, because the natural frequency of the suspension is commonly as low as 1 cycle per second. This brings the critical speed down to 26-1/2 miles per hour. Obviously damping must be provided to take care of this critical condition but at high speeds, where a great deal of the operation takes

place, very little, if any, damping is required. In fact, damping in this operating range does more harm than good, because it adds materially to the shock transmission. Therefore, it is important to keep the damping to the minimum amount required to prevent build-up of oscillation at the critical speed.

The commonest type of excitation of road vehicles consists of random transient impulses that are not regularly spaced. Occasionally, the road contour may approach a continuous sine wave of a few cycles duration. Nevertheless, the possible amplitudes are large enough to call for a fairly high, and carefully calibrated degree of control to avoid sustained oscillation, and bottoming of the suspension, except under the occasional very extreme condition. In fact, the job of arriving at an optimum balance between control of body oscillation and minimizing of harshness due to shock transmission, is indeed a delicate one.

Damping Devices Don't Absorb Shock

When we turn to the subject of damping devices we find ourselves confronted with one of the most glaring misnomers in the whole art of automotive engineering, namely, the term "shock absorber". Now, of course, a so-called "shock absorber" is nothing of the kind, but purely and simply an energy absorber. The only true shock absorber we have in a suspension is the spring itself. The sad truth is that most of the shock transmission takes place in the "shock absorber" because it tends to by-pass the spring, when the system is subjected to high velocity excitation. This is especially so in the case of a hydraulic device, which paradoxically is the one type of damper that is almost invariably described as a "shock absorber".

The need for damping arises from the fact that spring materials have negligible inherent energy absorption, so that once set into oscillation, a coil or bar can go on endlessly, alternately storing and discharging kinetic energy. The only way in which we can limit the oscillation or cause it to decay is to introduce external resistance to the motion and to dissipate the resulting work in the form of heat.

Damping devices can be divided broadly into two principle classifications, depending on whether they absorb energy by overcoming fluid friction or dry friction. In either case, wide variations in characteristic are possible in practice but for analytical purposes it is convenient, and therefore customary, to assume true viscous friction in one case and constant friction in the other. The law of viscous friction is that the resistance is a linear function of velocty of fluid flow. The assumption of constant dry friction has always been an academic ideal, and named "coulomb" after the great early scientist who first propounded it. By definition coulomb friction is constant in magnitude, independent of velocity or amplitude and always acts to oppose the motion.

Although damping is introduced into a suspension system primarily to control the amplitude of motion, it is of the utmost importance also to evaluate the penalty that must be paid in the form of increased force transmission. This is especially true of viscous damping, yet strangely enough, the force factor is too often lost sight of in the preoccupation with the familiar characteristic curves of amplitude ratio vs. frequency. It is for this reason that this paper stresses the force transmission characteristics in analyzing and comparing viscous and coulomb damping.

III. COMPARATIVE ANALYSIS OF VISCOUS AND COULOMB DAMPING (5)

The following analysis is primarily devoted to vibrating systems of one degree of freedom for two principal reasons:

(1) The characteristics of coulomb damping have not been fully explored as applied to more complex systems, and
(2) Most practical applications of coulomb damping fall within this category.

However, the specific case of automobile suspension is examined to show why viscous damping is inherently better adapted than the coulomb type to the essential requirements of such a system.

A - Free Vibration

In all treatments of the subject one characteristic of coulomb damping is readily demonstrated, namely, that the reduction in amplitude of free vibration in successive cycles is a constant quantity,

$$\Delta X = \frac{4F}{K}$$

where ΔX = reduction in amplitude per cycle
F = the constant friction force
K = the spring rate.

Therefore, the damping effect is independent of the vibrating mass as well as of the magnitude of the displacement.

In contrast, viscous damping of a free vibrating system is characterized by a constant amplitude ratio from cycle to cycle of the form

$$\frac{x_m}{x_m + 1} = e^{\frac{2\ b}{\sqrt{1 - b^2}}}$$

where b = damping factor, relative to the critical value $C_{cr} = 2\sqrt{KM}$
$x_m, x_m + 1$ = displacement amplitude of two successive cycles.

The comparative effects of the two types of damping are best visualized by a force-displacement diagram, particularly because such a diagram graphically defines the integrated energy changes. As shown in Chart 3-A for coulomb damping, starting from initial amplitude of 1/2", the spring force and damping force are plotted separately against displacement. It will be noted that the ordinate scale is given in force units relative to K, the spring rate, and so the diagram will cover any case where the friction force in pounds is 1/10 of the spring rate in lbs. per inch. The resultant of the spring and damping forces is given by the resultant force line which, of course, determines the acceleration of the vibrating mass.

Since the area under the resultant force line represents the value of the integral $\int FdX$, it, therefore, defines the work done. Since the vibrating mass starts from rest and again comes to rest after each half cycle, the positive work of acceleration must be equal to the negative work of deceleration, represented by the shaded areas above and below the base line. The same condition must be met on the return motion so that at the end of one cycle the vibrating mass has come to rest at a displacement of 1/10" (point 4) or at a reduction in amplitude of 4/10", equal to four times the given ratio of friction to spring rate. It will be noted, however, for the case represented in this chart, that the resultant force at point 4 has become zero because the spring restoring force and the friction force are equal. Therefore, the free vibration is completely damped in one cycle for the conditions given.

A similar force diagram for viscous damping is given in Chart 3-B for the same initial displacement and the same final displacement after one cycle. It will be noted that the damping force is now variable and the curve shown was developed point for point from the rationally derived equation of motion. The resultant force curve, as in the previous chart, meets the condition that the work during acceleration equals the work during deceleration, as indicated by the corresponding shaded areas. However, in contrast to the coulomb damped vibration, the resultant force at the extreme displacements

is necessarily identical with the spring restoring force since the damping force falls to zero at zero velocity. Therefore, at point 3 after one complete cycle of oscillation, it will be noted that the mass has only momentarily come to rest. Oscillation will continue with a constant ratio of reduction in amplitude per cycle but, theoretically, it is plain that the body would never come to rest under true viscous damping. Of course, in any practical case there is always some extraneous friction force which is actually relied on to completely damp the motion.

It will be observed that the maximum decelerating force acting on the mass occurs at a point short of the maximum displacement. This is an important characteristic of viscous damping inasmuch as the damping force is necessarily out of phase with the displacement. Also, it is the basic reason why the resultant force rather than the displacement amplitude is the really significant factor in viscous damping.

An interesting side light on this comparison of viscous and coulomb damping is its interpretation in terms of human reaction in the case of vehicle suspension damping. As brought out in the discussion of Chart 1, for the frequency range up to six cycles per second, human tolerance of harmonic vibration is directly related to the maximum rate of change of acceleration, or "jerk". It might appear that the abrupt change in acceleration with coulomb damping at the extreme displacements, as a result of the reversal in the direction of the friction force, would contribute seriously to passenger discomfort. However, the fact is that the disturbing effect of high jerk values is only associated with appreciable duration, as is indicated by the low frequency range of harmonic motion in which it is important. Since the high jerk rate has zero duration in the case of coulomb damping, it should have no discernible effect. This conclusion is in keeping with actual observations in vehicles having suspension systems of one degree of freedom.

B - Steady State Forced Vibration

The case of steady state forced vibration presents a very different picture from that of free vibration and brings into much sharper contrast the diverse characteristics of coulomb and viscous damping.

Under steady state forced vibration the transmitted forces through a simple vibrating system are well established both for the undamped condition and for viscous damping. However, as pointed out earlier, the case of coulomb damping does not lend itself readily to straightforward mathematical treatment. Nevertheless, Den Hartog (2) has worked out an ingenious approximation based on equivalent viscous damping which should be valid at least in evaluating the forced vibration amplitude. His analysis also brings out very clearly an important limitation of coulomb damping, namely, that it will not restrict the amplitude of resonant vibration unless the friction force is greater than $\frac{\pi}{4} P_o$, where P_0 = the maximum amplitude of the driving force. This can be seen from the fact that the work done by a harmonic driving force at resonance is equal to $\frac{\pi}{4} P_o X$, whereas the work done by the constant friction force, F, is equal to FX. Consequently, if F is less than $\frac{\pi}{4} P_o$, the amplitude will increase on successive cycles theoretically to infinity. It is evident, then, that coulomb damping is not applicable to a system which is subjected to steady state vibration at resonance unless the maximum value of the driving force is definitely limited and at a known level.

In connection with viscous damping characteristics, the treatment given by N. O. Myklestad (3) has been closely followed. This is exceptionally clear in differentiating between amplitude magnification and force transmissibility.

The comparison of viscous and friction damping effects in steady state forced vibration has been made for the following specific conditions:

(a) Driving force proportional to the square of frequency,
$P_n = P_0 r^2$
where P_0 = maximum driving force at resonance
and $r = \frac{\omega}{\omega_n}$
ω = driving force frequency
ω_n = natural frequency of system
This condition was selected because it is characteristic (1) of a vehicle traversing a fixed irregularity at variable speed; and (2), of the unbalanced forces in all variable speed machines.

(b) Coulomb damping friction, $F = P_0$

(c) Viscous damping at 25% of the critical value. This gives substantially the same maximum force transmission as the specified coulomb damping in the resonant frequency range.

Referring to Chart 4, the force transmission characteristics for both conditions of damping as well as for the undamped system are given for the frequency range up to $\frac{\omega}{\omega_n} = 7$. The ordinate scale is in terms of the ratio of maximum force transmission to maximum driving force at resonance, $(\frac{\omega}{\omega_n} = 1)$.

It will be seen in Chart 4 that the undamped condition, although capable of infinite amplitude at resonance, gives the lowest force transmission at high frequency ratios, approaching P_0 as a limit. With viscous damping at 25% of critical, the resonant amplitude is evidently well controlled but the force transmission at high frequencies increases continuously above a frequency ratio of 2 and begins to exceed the transmitted force at resonance at a frequency ratio of approximately 3.5. With coulomb damping at the specified level the maximum force transmission is substantially the same as with viscous damping near resonance but thereafter falls off continuously as the driving frequency increases, approaching a force transmission of $2 \times P_0$ as a limit. Thus, at the high frequency ratios coulomb damping adds to the force transmission only by the amount of the constant friction force.

It should be pointed out that, with friction equal to P_0, the friction force exceeds the maximum value of driving force, at all frequency ratios less than 1. Therefore, no motion of the vibrating mass is induced (see Chart 3) and the driving force is transmitted directly through the friction damper. Nevertheless, it will be seen that this force transmission is less than that obtained with either no damping or viscous damping in this frequency range. This follows from the fact that in either of the two latter cases the induced vibration introduces spring deflection forces which augment the resultant force transmission.

In contrast to the force transmission curves of Chart 4, Chart 5 shows the comparative displacement amplitudes of steady vibration under the same set of conditions. Note that the ordinate scale is expressed in amplitude ratio with respect to the static deflection, P_0/K, corresponding to the maximum driving force, P_0, at resonance. As in Chart 4, the maximum harmonic driving force is assumed to vary as the square of the frequency ratio, so that

$$P_n = P_0 (\frac{\omega}{\omega_n})^2.$$

It will be observed that, at high frequency ratios above 3, there is no difference in the displacement amplitude induced either with coulomb or viscous damping, or with no damping. Consequently, the comparison on this basis gives not the slightest inkling of the greatly augmented force transmission that takes place through a viscous damper in this frequency range.

Chart 5 also shows that the peak amplitude near the resonant frequency is 30% lower with specified amount of coulomb damping than with 25% of critical viscous damping. This result must be qualified by the limitation, previously emphasized, that the damping friction must exceed $\frac{\pi}{4}$ times the maximum driving force at resonance to keep the amplitude from going out of control. The following conclusions can be drawn from these results with regard to steady state forced vibration of a system having a single degree of freedom:

1 - Where the operating conditions are confined to high frequency ratios the undamped condition is the most desirable with regard to force transmission.
2 - Where it is necessary to control resonance, coulomb damping is very definitely to be preferred as long as the maximum driving force is known.
3 - Viscous damping is desirable only where operating conditions require resonance control under driving forces of unknown magnitude.

C - Transient Forced Vibration

The important practical condition of excitation by a single impulse is particularly difficult to evaluate by long hand, analytical methods. Consequently, the problem has been programmed on a digital computer and calculations are now in progress to determine the force transmission under each of the damping conditions investigated for steady state vibration.

Preliminary results indicate that the relative magnitude of force transmission under transient excitation is very similar to the comparative results shown for the steady state at high ratios of natural vibration period to impulse duration. The comparable range of steady state excitation would be at frequency ratios, $(\frac{\omega}{\omega_n})$, greater than 3:1.

Impacts are commonly encountered in vehicle operation, corresponding to high amplitude impulses of short duration. Under these conditions, the shock transmission with viscous damping, even when limited by blow-off valves, can build up to serious proportions.

D - Effect of Variable Mass With Fixed Damping

As the previous discussion has shown, coulomb damping is always independent of the sprung mass, either in free or forced vibration. This characteristic makes it ideal for application to suspension systems which are subject to variable load. An excellent case in point is damping of railroad freight truck suspensions where the range of sprung load from full capacity to light car may be of the order of 5 to 1.

In the case of viscous damping, the relative damping factor depends upon the ratio of the absolute damping constant to the critical value. In a simple system the latter is defined as $C_{cr} = 2\sqrt{KM}$. Therefore, with a fixed spring rate and a viscous damper having a fixed damping constant, a variation in mass results in a change in the relative damping factor. This may be expressed as follows:

$$b = \text{damping factor} = \frac{C}{C_{cr}} = \frac{C}{2\sqrt{KM}}$$

$$\text{If C and K are constants, then } \frac{b_1}{b_2} = \sqrt{\frac{M_2}{M_1}}$$

This relationship is shown graphically in Chart 6 by the curve designated "percent of critical damping" and it will be seen that with a nominal damping factor of 25% at a relative mass of unity, the damping factor will increase to 35% when the mass is reduced one-half and will decrease to 18% when the mass is doubled. The corresponding amplitude ratios in free vibration, also shown in Chart 6, would be, respectively, .10 and .32 as compared with the value of .2, at the nominal 25% damping factor.

E - Automobile Suspension Damping

Automobile suspensions, reduced to their simplest terms, are essentially systems of two degrees of freedom, considering each wheel as part of a separate system. This consists of a portion of the sprung mass, M, the suspension spring of rate K, the wheel and associated unsprung mass, m, and the tire having a spring rate, k. Their relative magnitudes are commonly in the ratios

$$k/K = 7 \text{ to } 9 \text{ and } \frac{M}{m} = 6 \text{ to } 10.$$

The problem is to provide adequate damping of each of the two masses with a single interconnecting device. Since the natural vibration frequencies of the two masses are widely different, the damping requirements for each mass can be considered independently, as though it were part of a separate single degree of freedom system. Accordingly, the unsprung mass, m, is acted upon by the tire and spring in parallel having combined rate K + k, while the unsprung mass can oscillate on the supporting spring of rate K. Detailed calculations are given in Appendix 2 for a typical case, comparing the coulomb and viscous damping requirements for both masses.

Considered in general terms, if coulomb damping were to be used, the amplitude reduction per cycle in free vibration would be $\frac{4F}{K+k}$ for the unsprung and $\frac{4F}{K}$ for the sprung mass. Since (K + k) is of the order of 8 to 10 times K, a given friction setting would produce only 1/8 to 1/10 as much damping effect on the unsprung mass as would be exerted on the sprung mass. Referring to the calculations of Appendix 2, this is quite evidently out of proportion to the possible amplitudes of motion.

On the other hand, the viscous damping factor is related to the product of spring rate and mass, $(b = \frac{C}{2\sqrt{KM}})$ so that, if C is a constant, corresponding to a common damper:

$$\frac{b_s}{b_u} = \sqrt{\frac{(K+k)\,m}{K\,M}}$$

where b_s = damping factor for sprung mass, Pct. of critical
b_u = damping factor for unsprung mass, Pct. of critical

It is apparent then that the viscous damping constant will be of the same order for both masses, since the high combined spring rate and small unsprung mass have a product closely equivalent to that of the low spring rate and high sprung mass. Consequently the calculations show that a viscous damper can be calibrated to provide comparable damping requirements for both body motion and wheel hop.

IV. ACTUAL CHARACTERISTICS OF HYDRAULIC AND FRICTION DAMPERS (5)

As indicated in the introduction, dry friction dampers have fallen into bad repute largely because of the materials used for the friction surfaces. It is characteristic of practically all dry rubbing friction between metallic surfaces that the static and dynamic friction coefficients are widely different. In general, the starting or breakaway resistance is high and falls rapidly as the velocity increases. The obvious result is that the damper has a high impact transmission factor in relation to the energy absorbed during relative motion of the friction surfaces. Referring again to Chart 1-A, it is obvious that variable friction of this character, even in free vibration, would seriously increase the maximum acceleration for a given energy absorption.

Fortunately, it has been found that certain special non-metallic molded materials, of the type employed in brake lining, in contact with steel surfaces approach very closely to a uniform coefficient of friction, with the static and dynamic coefficients substantially identical. This means that the ideal coulomb type of damper can be made a reality. Furthermore,

materials of this type have a much lower rate of wear and are insensitive to temperature, moisture conditions, etc.

The nature of the variation in friction coefficient with velocity between any two materials can be strikingly demonstrated by a device illustrated in Fig. 1. This consists of two drums geared together and driven counter to each other at the same speed, with the left-hand drum rotating in the clockwise direction. A bar placed across the two drums and given an initial velocity will obviously be forced to move at different relative velocities with respect to each of the two drums. If the coefficient of friction is independent of velocity the action of the bar will be stable. Once disturbed from its center position it will oscillate at the same or gradually diminishing amplitude. However, if the coefficient decreases with increasing velocity, the action of the bar will be unstable and a self-excited vibration will be induced until the bar either leaves the rolls or is otherwise restrained. It will be noted that the drums have two adjacent grooves so that the action of any two materials can be compared simultaneously. When a bar lined with the special composition material is placed on the drum side by side with an unlined steel bar the stability of the lined bar and the instability of the unlined bar are immediately apparent. In fact, the unlined steel bar is self-exciting without external disturbance. It is of interest to note that this device is described by Timoshenko (4) as a means of experimentally determining the coefficient of friction, on the assumption that it remains constant. However, he does not mention the consequences of variability in the coefficient with velocity.

The mechanism of the action can be readily seen when it is considered that the friction driving force acting on the bar at each drum is a product of the weight reaction on the drum and the coefficient of friction. When the bar is displaced from its center position the friction force increases at the drum carrying the greater weight reaction and decreases at the drum carrying the smaller reaction. This force differential tends to restore the bar to its normal position, and it will be directly proportional to the displacement of the bar only if the coefficient of friction is the same at both drums. This will clearly be true when the bar is not moving. However, if the bar is in motion its relative velocity with respect to the two drums will be different, and, therefore, the linear relation between restoring force and displacement will hold only if the coefficient of friction is independent of velocity. Stated another way, the static and dynamic coefficients must be identical.

Now let us examine the consequences if the friction coefficient should decrease as the velocity increases, a condition characteristic of dry metal to metal friction. Since the relative velocity of the moving bar is greater with respect to the drum it is approaching, the coefficient of friction will be lower at that drum than at the one toward which the bar was initially displaced. This means that the force available to accelerate the bar at a given displacement will be greater than the force acting to decelerate the bar at the same displacement in the opposite direction from the center position. Since the work of deceleration must equal the work done during acceleration before the bar can be brought to rest, it follows that the final displacement in the opposite direction will be greater than the initial displacement. In this way, the amplitude of the bar increases on each succeeding pass across the rolls until it either hits a stop (as in the model) or falls off the rolls.

On the other hand, if the coefficient of friction is independent of velocity, the restoring force will be equal for the same displacement in either direction and the initial and final displacement amplitudes will also be equal. Theoretically, therefore, the bar should continue to oscillate at a uniform amplitude once disturbed. However, with the special constant friction lining material it is found that the amplitude will gradually diminish, indicating that there is actually a slight increase in coefficient of friction with increase in relative velocity of motion.

At the left hand side of Chart 7 is shown an actual indicator card obtained from the Chrysler Design constant friction snubber employing the special composition material operating against a soft steel barrel. This record was obtained on a machine which is positively driven with very nearly harmonic action. It will be noted that the card is substantially rectangular

and very nearly symmetrical about the base line, indicating practically constant friction throughout the stroke in both directions.

It is of interest to compare this record with that on the right hand side of Chart 7, obtained on the Chrysler "Oriflow" type viscous damper, which probably comes closer than any other device to the true viscous damping characteristic. This is brought out by the superimposed diagram which represents the theoretical variation in resistance in proportion to velocity, required to give the same energy absorption as that actually obtained. The superimposed dashed rectangular diagram represents the equivalent energy absorption with coulomb damping. Figure 2 shows a sectional view of the "Oriflow" unit, together with exploded views of its component parts.

Figure 3 is a cutaway view of the constant friction snubber as applied to railroad cars, and Figure 4 in the same photograph illustrates the similar unit adapted for tank installation. The internal construction is practically identical in both units and only the length of travel and external connections are different to suit the particular installation conditions. It will be seen in Figure 3 that the friction elements consist of three segmental shoes with the special lining cyclewelded on the outer surfaces. The spring-loaded flat 45° wedge surfaces on the plunger head and free wedge ring engage mating surfaces on the shoes and thus apply uniform radial pressure to the shoes against the cylinder barrel. The lower wedge ring is necessarily free to slide on the plunger and thereby, provides automatic take-up for wear. It will be noted that the shoes remain in fixed relation to the plunger and spring during operation of the snubber. Since the frictional resistance is approximately 60% of the spring pressure, the force required to compress the snubber cannot produce additional deflection of the spring.

Chart 8 illustrates the insensitivity of the new constant friction snubber to variations in velocity and temperature. The points plotted in the upper chart show the average friction obtained in tests made at different strokes and at different driven speeds. The scale of the abscissa is the product of stroke and rpm of the test machine and thus represents relative mean velocity. The ordinates show the average recorded frictional resistance in both directions. It will be noted that the overall variation from the average friction is of the order of plus or minus 4%, over the entire relative velocity range from 20 to 300.

The lower curve shows that the total variation in frictional resistance between - 60° and + 160°F. was confined to a range of 100 lbs. at a normal setting of about 1400 lbs. In this connection it will be noted that viscous dampers are necessarily sensitive to temperature by virtue of change in the fluid viscosity even with the relatively low viscosity index of the special fluids employed.

V. PRACTICAL APPLICATIONS OF CONSTANT FRICTION SNUBBER

The constant friction snubber was originally developed by the Chrysler Dynamics Research Department as an integral part of the Chrysler Design railroad freight truck and has been in commercial use on this equipment since 1948. The results have been highly successful both as to performance and life.

Extensive testing has proved that true coulomb damping produces a tremendous reduction in transmitted shock as compared with friction devices subject to high breakaway resistance and lacking precise control of the friction calibration.

This can be illustrated by typical test results obtained on two railroad trucks having identical spring suspensions and loads where the only difference affecting the vertical shock level was the type of friction damping employed. The reference truck, shown in Figure 5, incorporated a built-in type of snubber using alloy cast iron shoes operated against hardened steel wear-plates. The comparison truck was the Chrysler Design equipped with the constant

friction snubber, as shown in Figure 6. Identical cars equipped with these trucks were operated in the same test train at speeds up to 90 mph over a course of approximately 120 miles. The vertical shocks were recorded on contact type accelerometers, each consisting of a bank of four elements calibrated to different acceleration intensities ranging from 25% g. to 75% g. Table 1 shows the recorded counts of each shock intensity in two representative runs with different loadings and spring groups but with these conditions identical in both trucks in each test. It will be seen that the number of vertical shocks was far lower in the truck with constant friction snubbers. Reductions range from 87.5% to 72% at the light load and amounted to 96% in the lower accelerations with the heavy load. It will be noted that in the latter case no shocks were recorded with the constant friction snubber above the level of 35% g., although a number of shocks above 50% were obtained with the reference trucks.

Within the limited range of static deflection possible in freight truck suspensions, the shock transmission is especially sensitive to the load. For example, under the test conditions of Table I the light load produced about 1" static deflection, and the heavy load about 2" static deflection. The striking reduction in the number of recorded shocks with the heavier load, even with the inferior damping of the conventional truck, again testifies to the fundamental fact that there is no substitute for static deflection in a suspension system.

It is evident that higher static deflections can be obtained by reducing the suspension stiffness at a given load. This constitutes a potent argument for the use of special low rate spring groups in freight cars that can be restricted to the handling of light loads of fragile commodities. This expedient has already been adopted by at least one important railroad which has equipped all of its fast merchandise freight cars with springs of reduced capacity and correspondingly lower rates in keeping with the light lading handled.

A more recent application of the constant friction snubber is its adoption by Ordnance for use on the latest medium M-48 tank, replacing the hydraulic dampers previously used. This decision was made after exhaustive tests by Ordnance development and field forces which conclusively established the superiority of the new friction unit. When both types of damper were calibrated for comparable control of hull oscillation, the following advantages were definitely established for the friction snubber:

1. Increased field service life in the ratio of at least four to one over the hydraulic.
2. Excellent stability of hull position after firing. This is vitally important to avoid compensating for changes in inclination of the gun platform between rounds.
3. Reduction of noise, measured inside a light track vehicle, up to 12 Decibels, with an average reduction of 8 to 10 DB. at three locations and at three different speeds.
4. Sizable overall decrease in hull vibration with maximum reduction of 75% in acceleration intensity at some frequencies and vehicle speeds.

The very pronounced reduction in noise and vibration is clearly in line with the predicted tendency of constant friction damping to minimize impact transmission. As shown in the analytical comparison with viscous damping, the latter tends to produce seriously augmented force transmission at high frequencies. This would tend to be exaggerated over the rough terrain encountered in cross-country tank operation.

SUMMARY & CONCLUSIONS

1. The effective static deflection of a simple suspension system determines its ability to absorb shock, as well as the natural frequency of vibration. The latter, in relation to the forced frequency of vibration determines the degree of isolation possible.

2. From the standpoint of passenger comfort, the natural frequency of the suspension is basic in determining the tolerable amplitude of the principal motions. This amplitude is inversely proportional to the cube of frequency and hence directly proportional to the 1.5 power of static deflection.

3. In a vehicle suspension system when the sprung mass has two principal degrees of freedom, namely, pitch and bounce, on two longitudinally spaced suspension elements, the latter should be located as close as possible to the centers of percussion of the sprung mass. If this condition is met the absolute wheelbase is of minor importance to the ride motion.

4. The damping requirements of a suspension system are determined by the nature of the excitation, particularly as it involves the possibility of resonance. Dampers absorb energy but transmit shock, and, therefore, the amount of control should be kept to the minimum required under critical resonant conditions.

5. Damping devices can be divided broadly into two types, according to whether they absorb energy by fluid or dry friction. The corresponding idealized characteristics are, respectively, viscous friction, in which the resistance to motion is proportional to velocity and coulomb friction, in which the resistance to motion is a constant force.

6. In a suspension system having a single degree of freedom, coulomb damping is generally preferrable to viscous damping because of its very much lower force transmission at high ratios of forcing frequency to natural frequency. The one qualification is that in steady state resonant vibration, the coulomb friction exceed 78.5% of the maximum sinusoidal driving force to control the vibration amplitude.

7. In automobile suspension systems, analysis shows that coulomb damping is not inherently well adapted to control both the sprung and unsprung masses. On the other hand, a single viscous damper can exert a comparable degree of control in both masses.

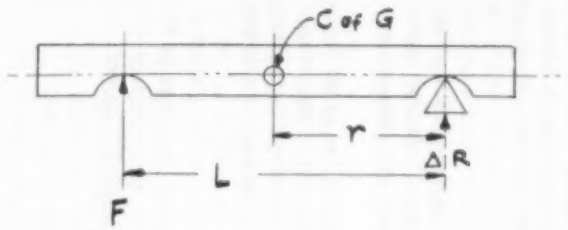
8. Coulomb damping possesses the inherent advantage over viscous damping of being independent of the mass of the system and related only to the spring rate. This characteristic is of great practical value in such vehicles as freight cars where the sprung load commonly varies over a range of five to one between the fully loaded and empty car.

9. A friction damper has been developed which achieves true coulomb damping by the use of friction materials which are unique in having the same static and dynamic coefficients. This unit is now giving highly satisfactory results in railroad freight car trucks and army track vehicles both as to performance and life.

BIBLIOGRAPHY

1. "Vehicle Vibration Limits to Fit the Passengers" by R. N. Janeway. SAE Journal, August, 1948, p. 48. Also, SAE Special Publication "Ride & Vibration Data" - p. 23.
2. J. P. DenHartog - "Mechanical Vibrations" (McGraw-Hill) p. 408.
3. N. O. Myklestad - "Vibration Analysis" (McGraw-Hill) p. 102.
4. Timoshenko - "Vibration Problems in Engineering."
5. Excerpt from paper on "Controlling Shock & Vibration by Friction Damping" by R. N. Janeway. Shock & Vibration Bulletin No. 21, Nov. 1953, (Confidential) Office of the Secretary of Defense - Research & Development.

ANGULAR ACCELERATION ASSUMING FIXED PIVOT



LET i = RADIUS OF GYRATION
M = MASS
α = ANGULAR ACCELERATION
F = APPLIED FORCE
ΔR = REACTION AT PIVOT

ROTATION ABOUT PIVOT, $FL = M(i^2 + r^2)\alpha$

ROTATION ABOUT C. OF G., $F(L - r) - \Delta Rr = Mi^2\alpha$

THEN
$$\frac{FL}{M(i^2 + r^2)} = \frac{F(L - r) - \Delta Rr}{Mi^2}$$

$$\frac{L}{i^2 + r^2} = \frac{(L - r) - \Delta Rr/F}{i^2}$$

$$\frac{\Delta R}{F} = \frac{r(L - r) - i^2}{i^2 + r^2}$$

FOR THE SPECIAL CASE WHERE THE CENTER OF GRAVITY IS EQUIDISTANT FROM THE SUPPORTS,

$r = L - r$ AND LETTING $i^2 = p\,r^2$

$$\boxed{\frac{\Delta R}{F} = \frac{1 - p}{1 + p}}$$

FOR $\Delta R = 0$, $r(L - r) - i^2 = 0$

$$L - r = \frac{i^2}{r} \quad \text{OR} \quad r(L - r) = i^2$$

AUTOMOBILE SUSPENSION DAMPING

1. VISCOUS

FOR CRITICAL DAMPING, $C_{cr} = 2\sqrt{KM}$ WHERE DAMPING FORCE $= C\dot{X}$

BODY OSCILLATION

AT ONE WHEEL, $K = 115$ LBS./IN. $W = 1150$ LBS. $M = \frac{1150}{386} = 2.98$

$$C_{cr} = 2\sqrt{2.98 \times 115}$$

$$= 37 \text{ LBS./IN./SEC.}$$

WHEEL OSCILLATION

K (TIRE) $= 800$ LBS./IN.

COMBINED SPRING RATE, $K = 800 + 115 = 915$ LBS./IN.

UNSPRUNG WEIGHT $= 123$ LBS.

UNSPRUNG MASS $= \frac{123}{386} = .32$

$$C_{cr} = 2\sqrt{.32 \times 915}$$

$$= 34.2$$

APPENDIX 2

AUTOMOBILE SUSPENSION DAMPING (CONTINUED)

TAKE $C = 7.0$, THEN

FOR WHEEL DAMPING, $b = \frac{7}{34.2} = .205$

AMPLITUDE RATIO $r = e^{-z}$ WHERE $z = \frac{2\pi b}{\sqrt{1 - b^2}}$

$r = e^{-1.315} = .269$ $z = 1.315$

FOR BODY DAMPING,

$b = \frac{7.0}{37} = .189$ $z = 1.21$

AMPLITUDE RATIO $r = e^{-1.21} = .30$

CONSIDER DAMPING OF FREE VIBRATION FROM INITIAL AMPLITUDE OF 2" FOR BODY AND .75" FOR WHEEL

THEN, AFTER 1 CYCLE, AMPLITUDE OF BODY = 2 x .3 = .60"

AMPLITUDE OF WHEEL = .75 x .269 = .20"

2. COULOMB DAMPING

FOR BODY DAMPING EQUIVALENT TO VISCOUS, $\Delta a = 2 - .60 = 1.4"$

$K = 115$

AMPLITUDE REDUCTION PER CYCLE, $\Delta a = \frac{4F}{K}$

$F = \frac{1.4 \times 115}{4} = 40.2$ LBS.

FOR WHEEL DAMPING EQUIVALENT TO VISCOUS, $\Delta a = .75 - .20 = .55"$

$K = 915$

$F = \frac{.55 \times 915}{4} = 126$ LBS.

RIDING COMFORT vs SPRING DEFLECTION

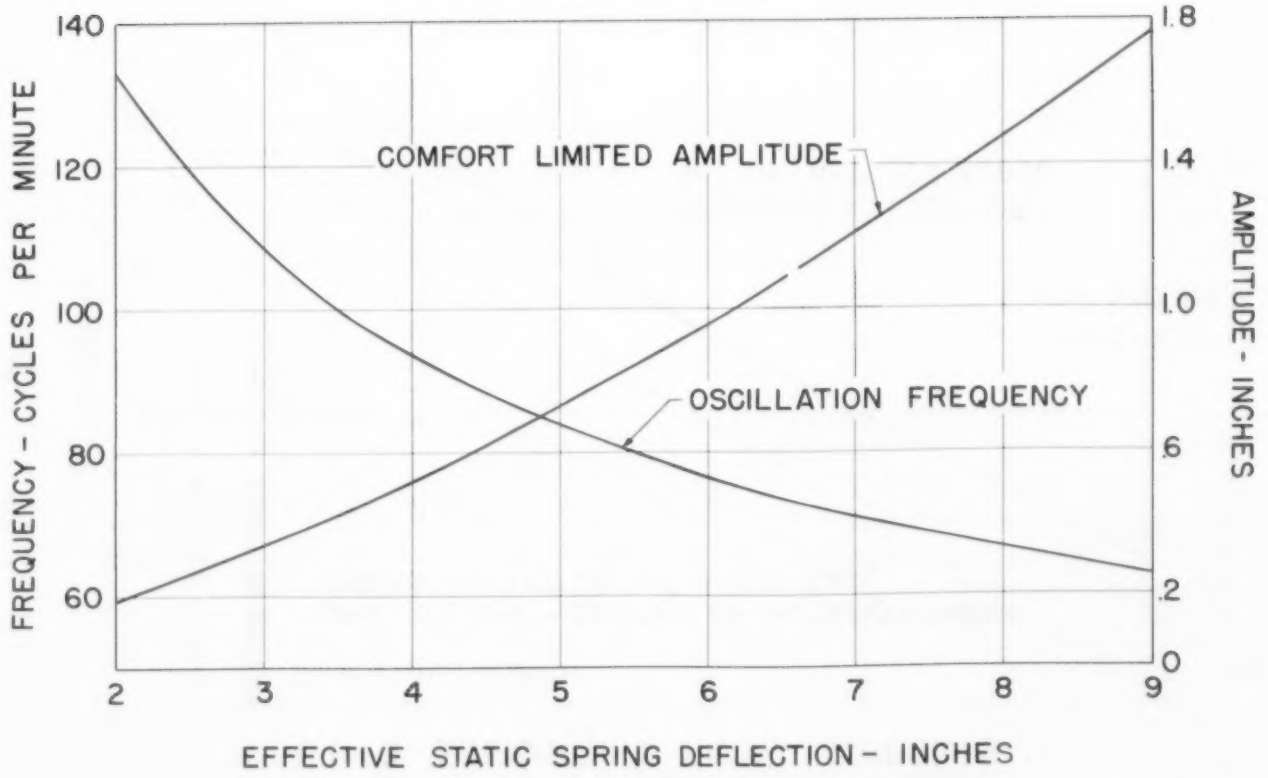


CHART I

EFFECTS OF MASS DISTRIBUTION ON REACTIONS DURING ANGULAR ACCELERATION C OF G EQUIDISTANT FROM SUPPORTS

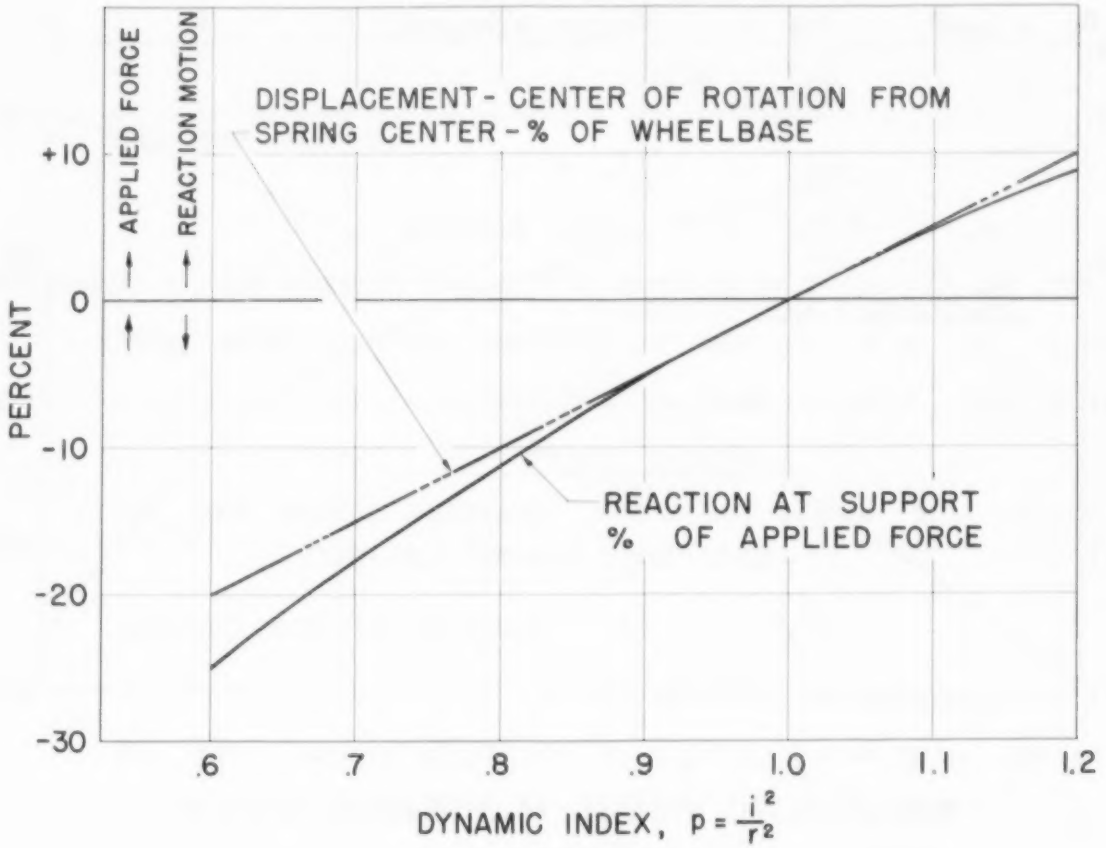


CHART 2

CHART 2

CHART 3A — FORCE AND ENERGY DIAGRAM

FREE VIBRATION WITH COULOMB DAMPING

CONDITIONS : INITIAL DEFLECTION (UPWARD) = .5 IN.
FRICTION LBS. = .10 x SPRING RATE (K)

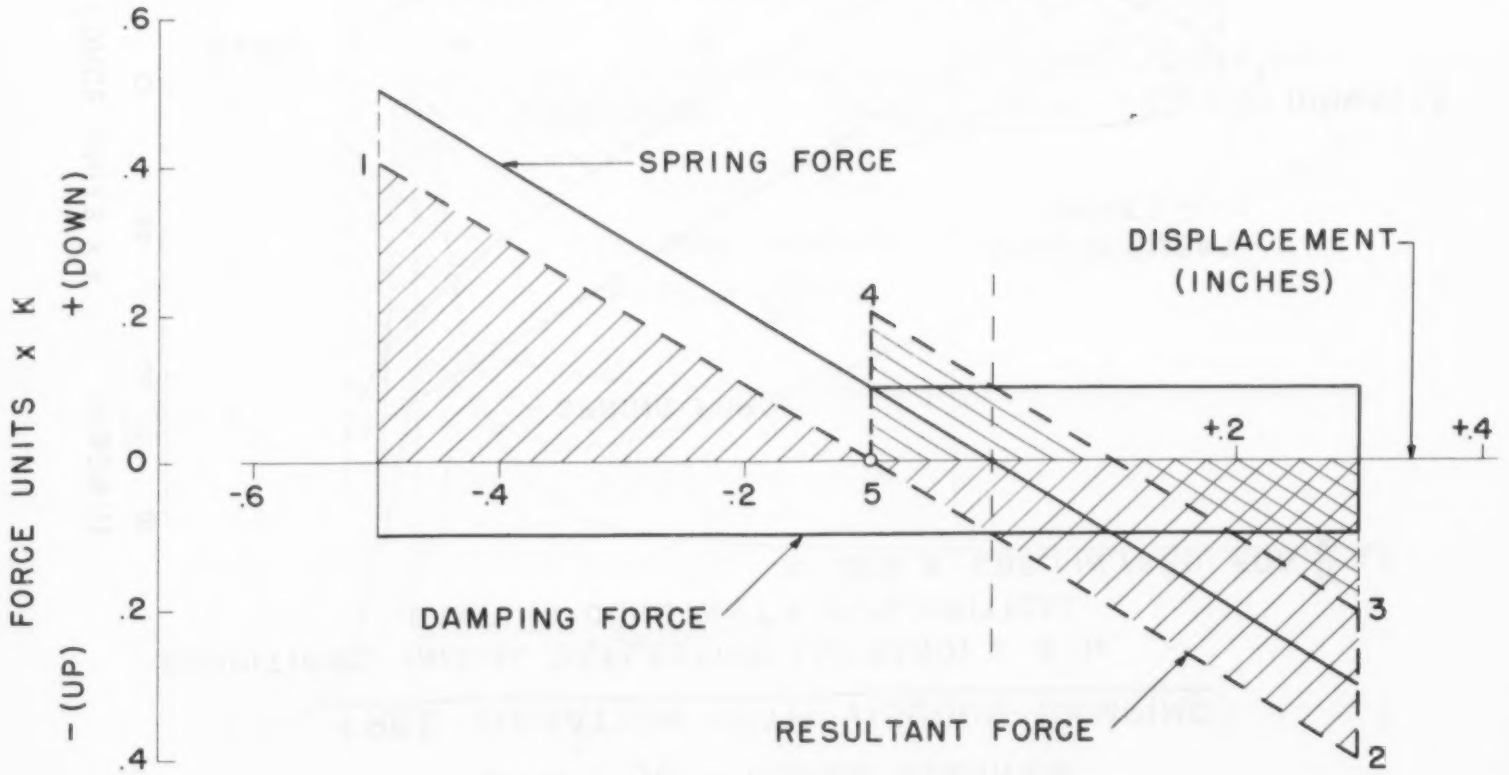


CHART 3B — FORCE DIAGRAM

FREE VIBRATION WITH VISCOUS DAMPING

CONDITIONS : INITIAL DEFLECTION (UPWARD) = .5 IN.
DAMPING CONSTANT = 25% CRITICAL
= .025 K LBS./IN./SEC. FOR $\frac{W}{K} = 1$

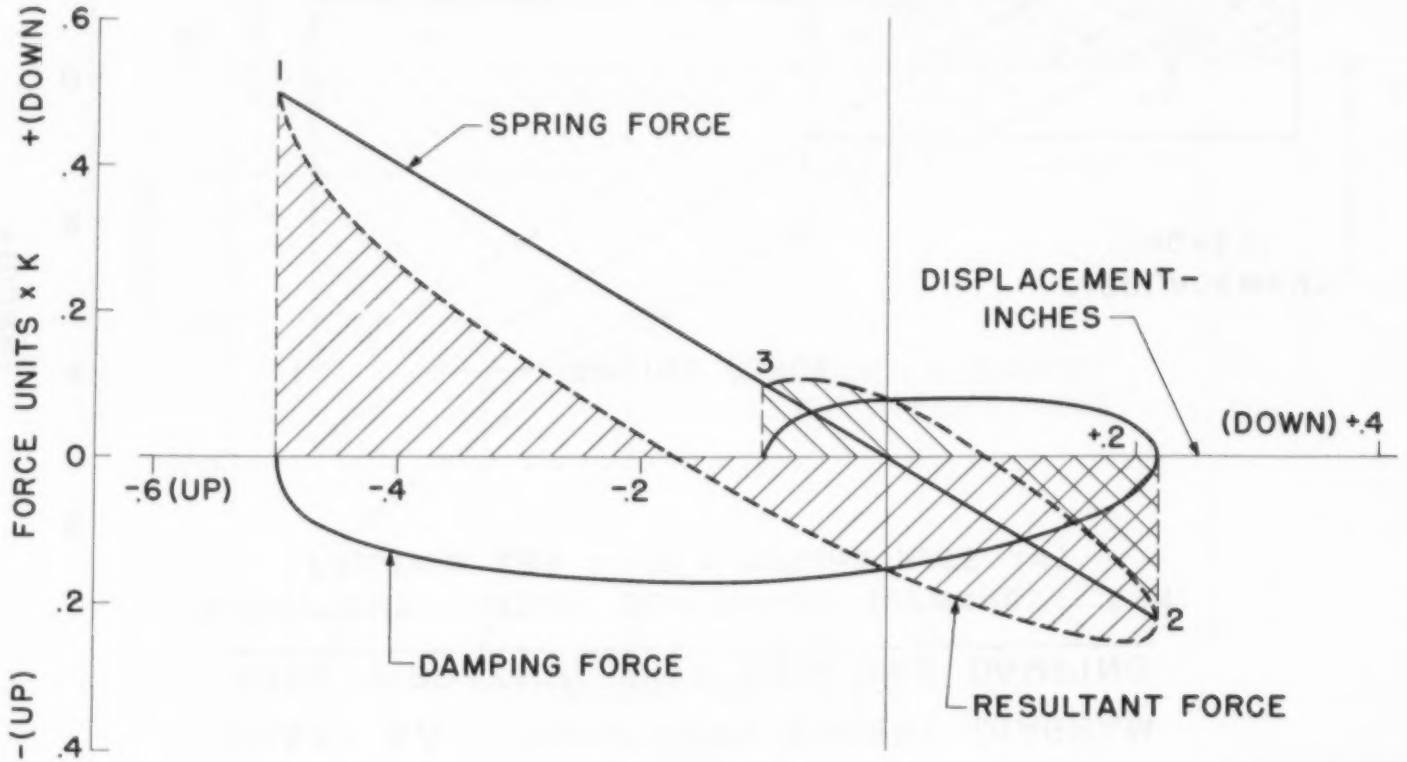


CHART 4

CHART 4
STEADY STATE VIBRATION
FORCE TRANSMISSION AT VARIABLE SPEED

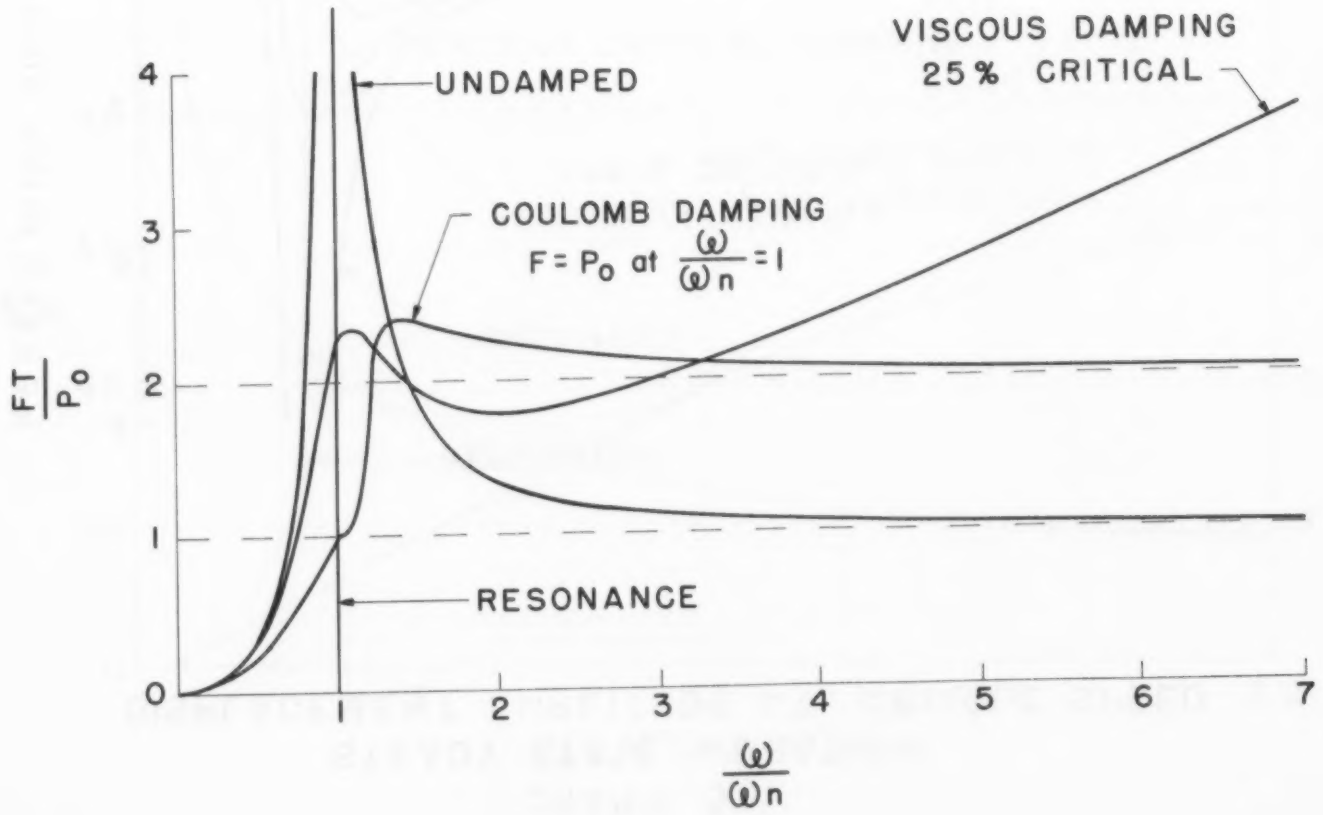
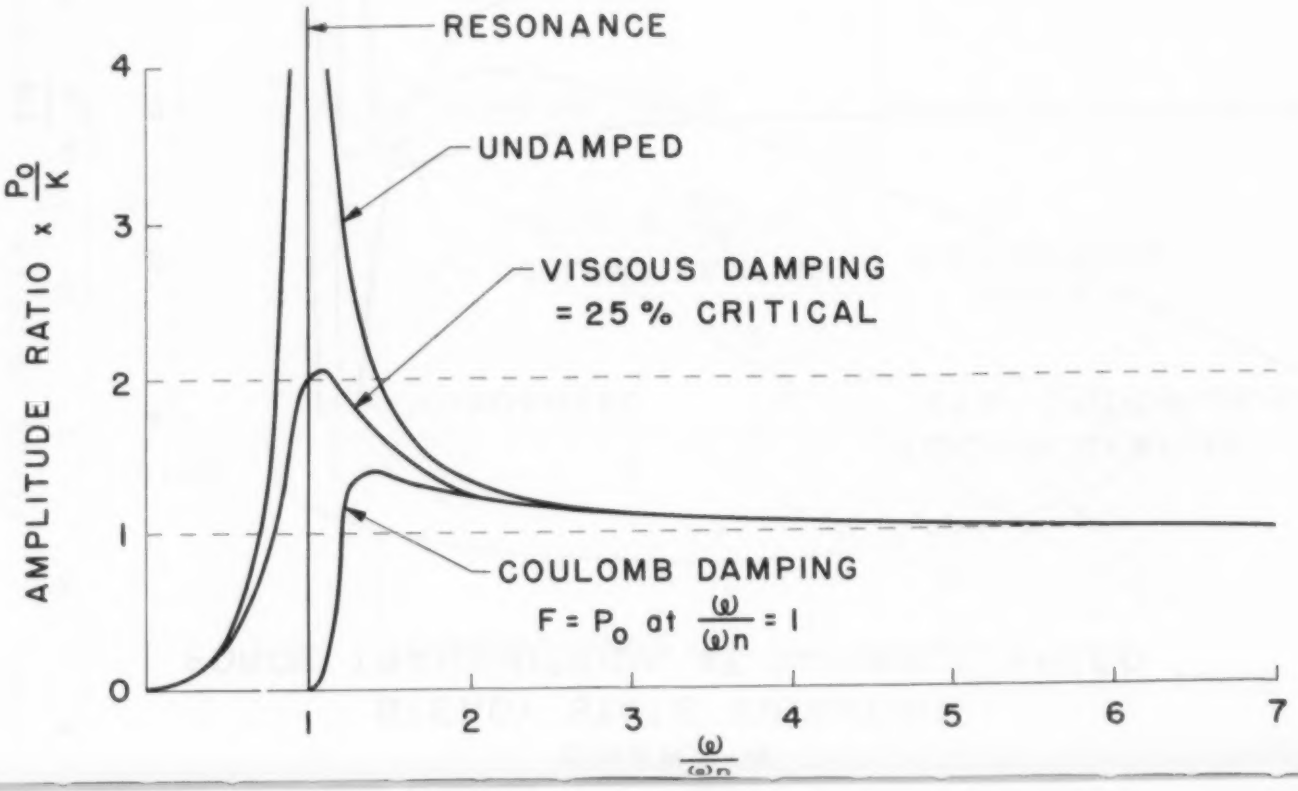


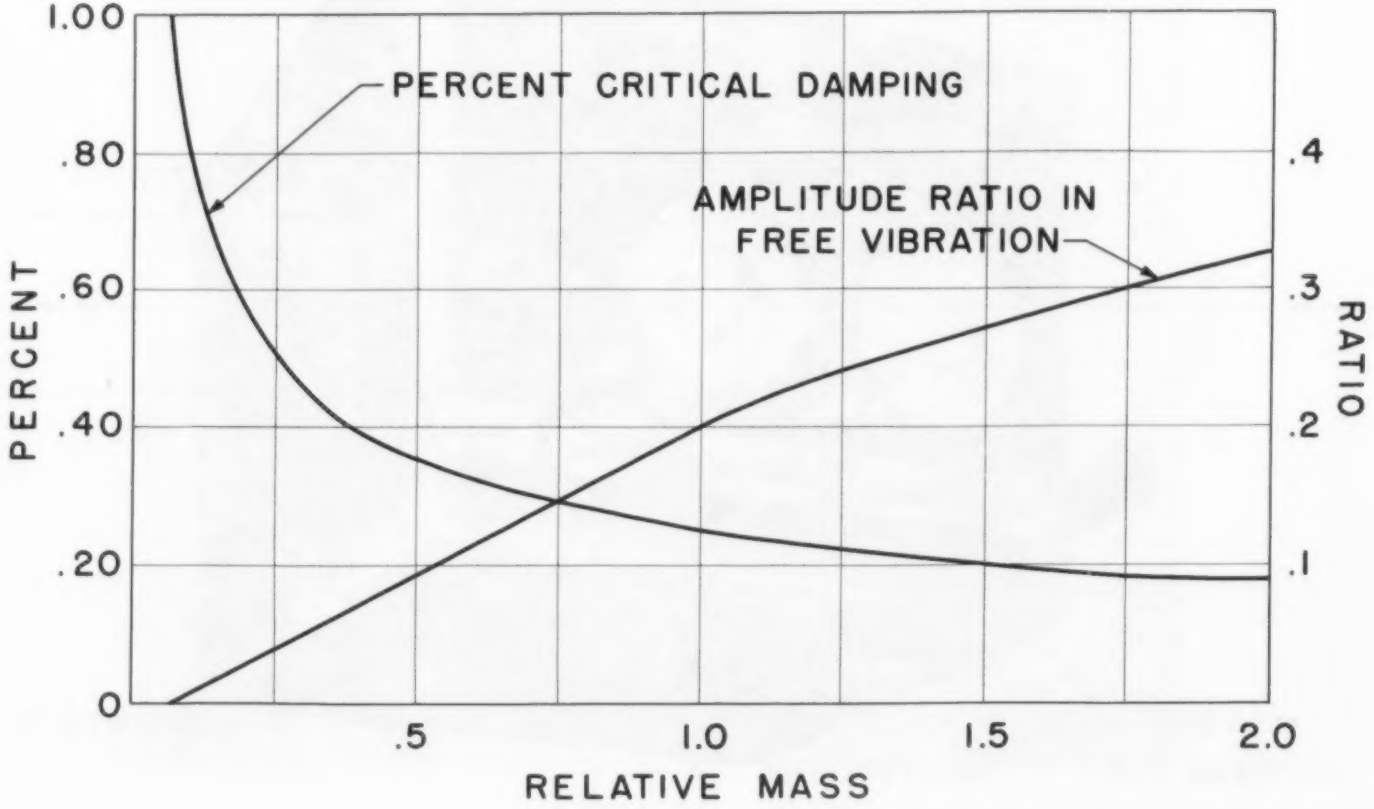
CHART 5
STEADY STATE VIBRATION
DISPLACEMENT AMPLITUDE AT VARIABLE SPEED



$\frac{\omega}{\omega_n}$

CHART NO. 6 — EFFECT OF VARIABLE MASS WITH CONSTANT VISCOUS DAMPING

25 % CRITICAL DAMPING AT RELATIVE MASS = 1



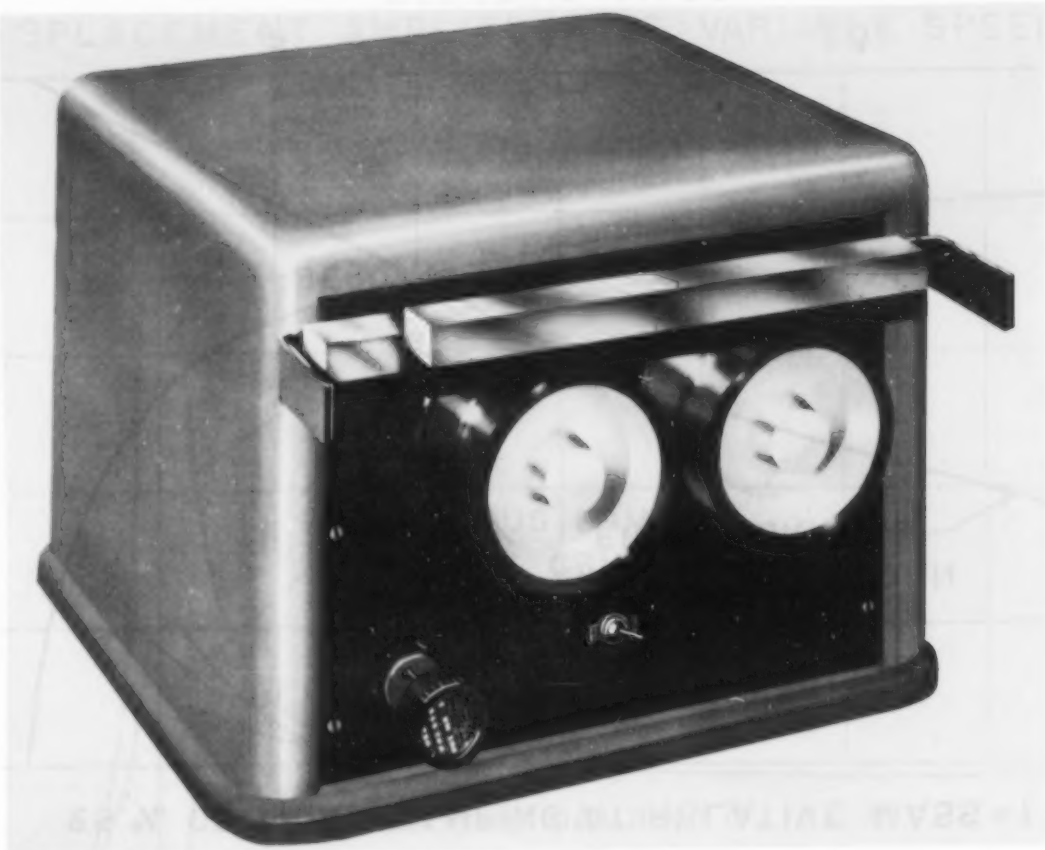
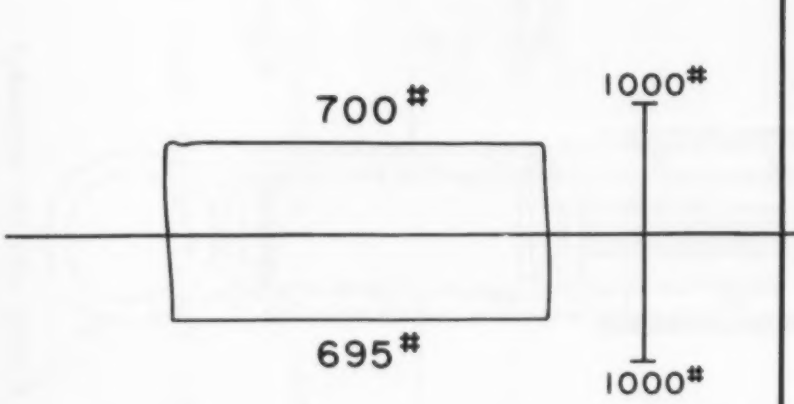


FIGURE I - DYNAMIC MODEL DEMONSTRATING
CONSTANT vs VARIABLE FRICTION COEFFICIENT WITH VELOCITY

CHART 7 - ACTUAL INDICATOR CARDS

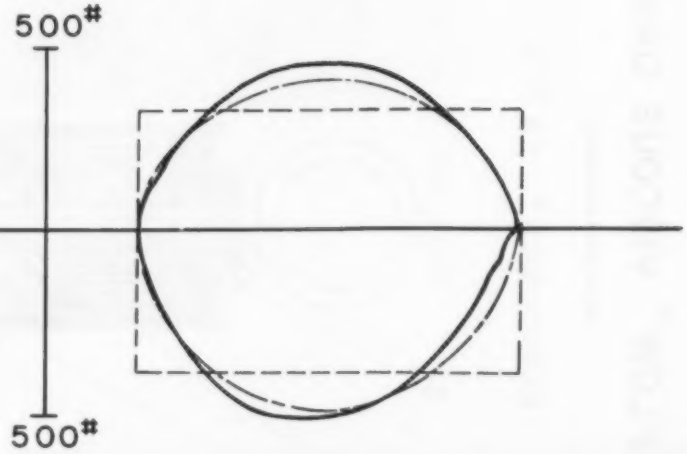
CHART 7 – ACTUAL INDICATOR CARDS

CONSTANT FRICTION SNUBBER



1" STROKE – 60 RPM

"ORIFLOW" HYDRAULIC DAMPER



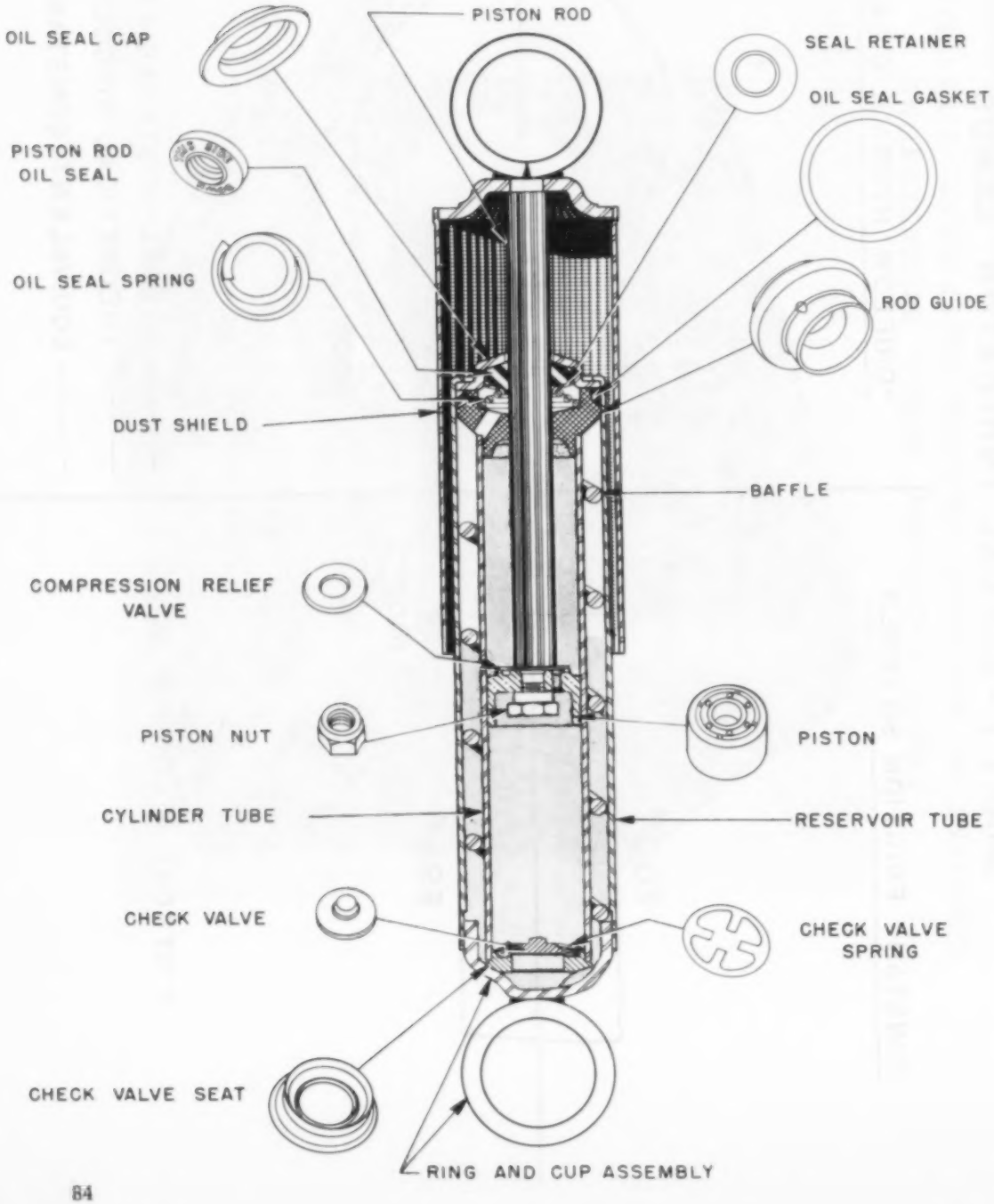
——— ACTUAL – 1" STROKE – 300 RPM

—·—·— THEORETICAL VISCOUS DAMPING

– – – – – EQUIVALENT CONSTANT FRICTION

FIG. 2
"ORIFLOW" VISCOUS DAMPER

COMPONENT PARTS



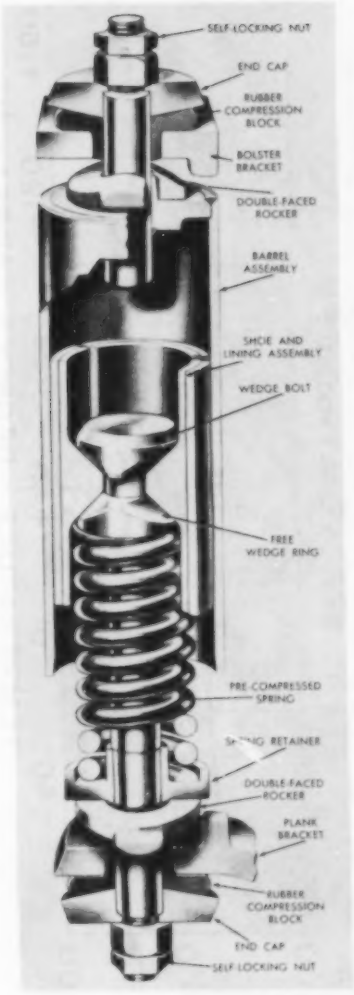


FIGURE 3–RAILROAD TRUCK UNIT
CUTAWAY VIEW

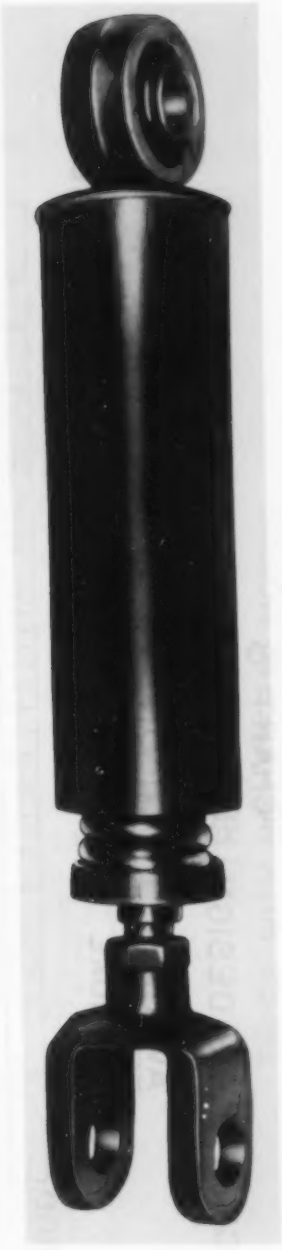


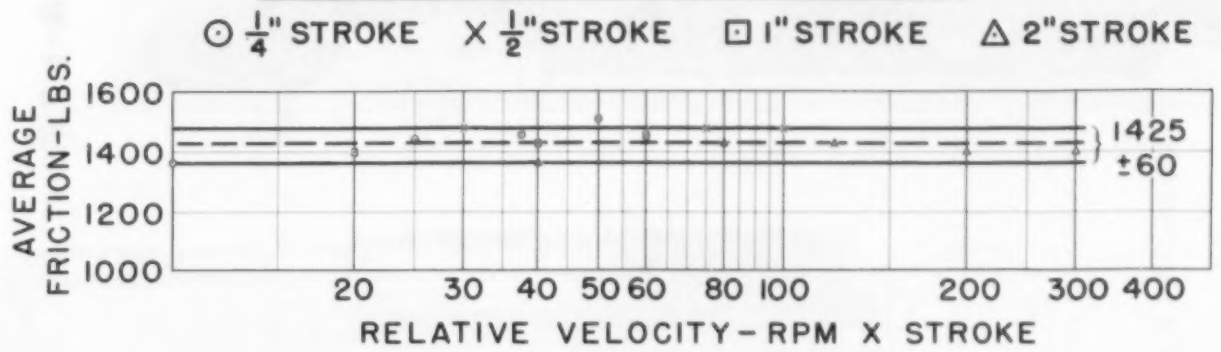
FIGURE 4–M48 TANK UNIT

CHRYSLER DESIGN CONSTANT FRICTION SNUBBERS

CHART 8

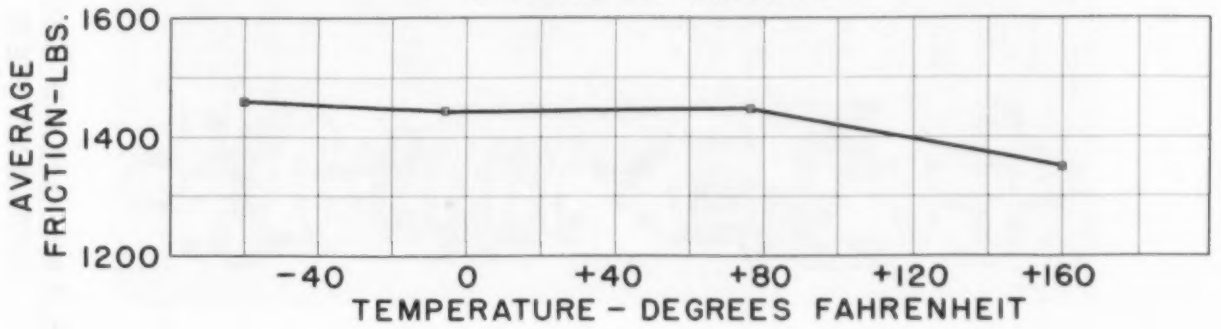
CHRYSLER DESIGN CONSTANT FRICTION SNUBBER

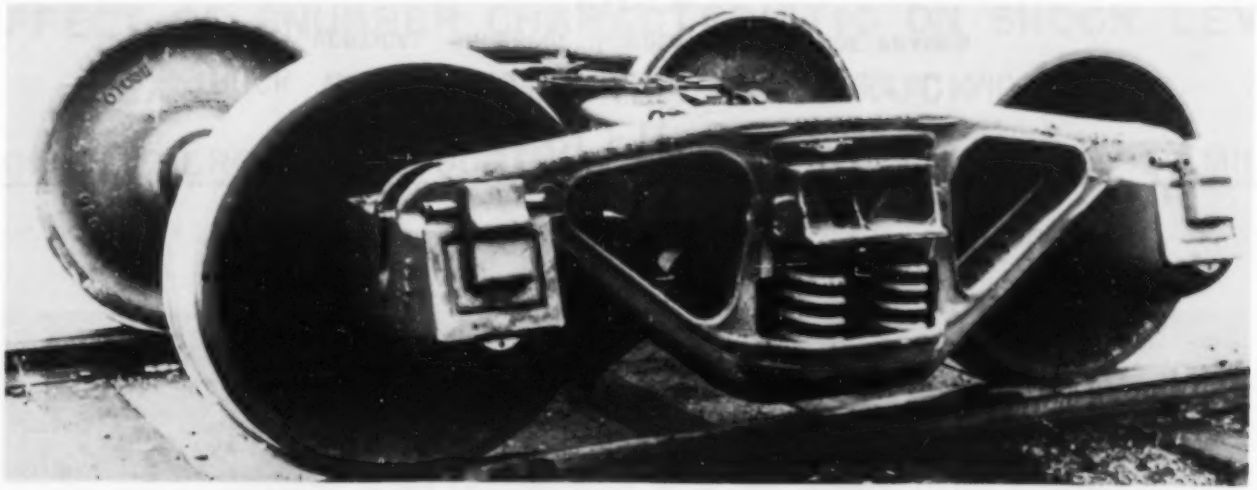
AVERAGE FRICTION VS VELOCITY

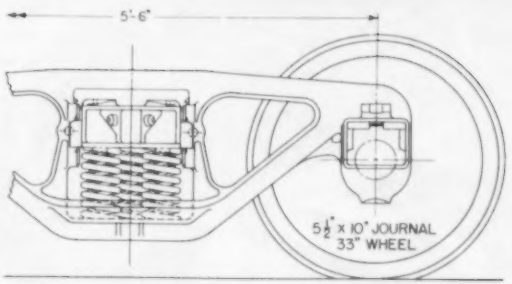


AVERAGE FRICTION VS TEMPERATURE

1" STROKE AT 60 RPM







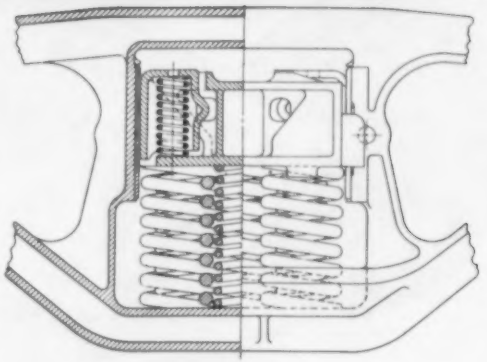
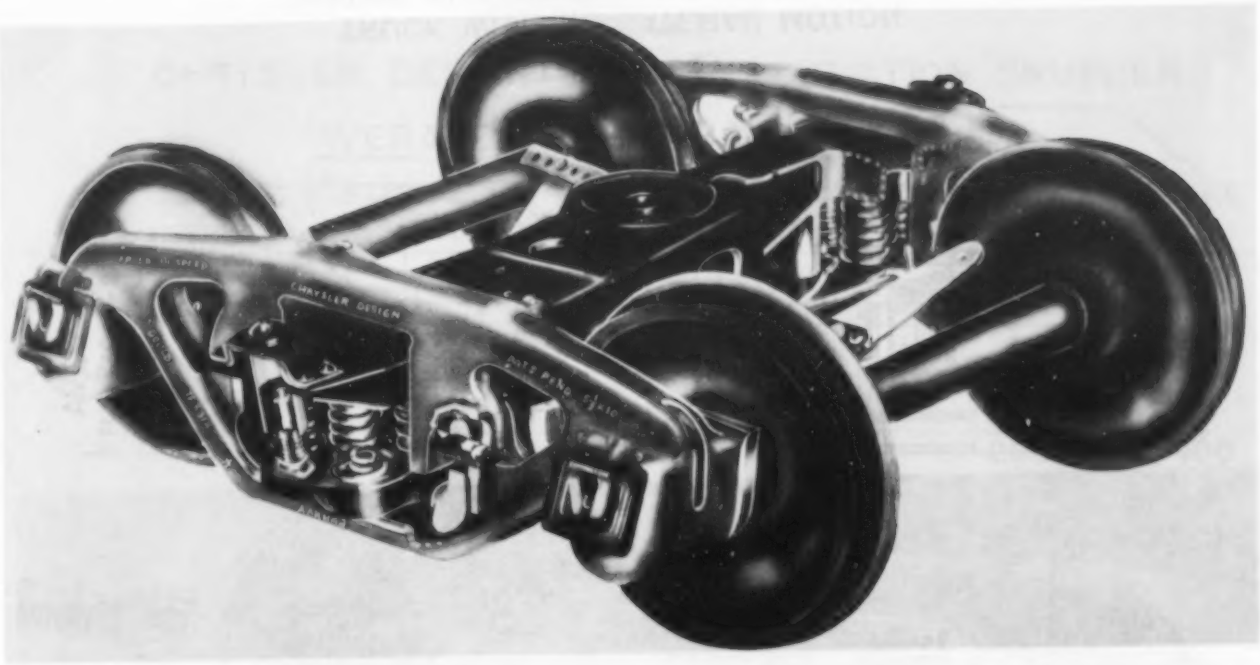


FIG. 5

TRUCK E

TRUCK WITH NO LATERAL MOTION

CONSTANT FRICTION SNUBBING ON COLUMN GUIDES



CHRYSLER DESIGN

TRUCK WITH SWING HANGERS FOR LATERAL MOTION

VERTICAL SNUBBING INDEPENDENT OF SIDE FRAMES

FIG. 6

TABLE I

TABLE I

EFFECT OF SNUBBER CHARACTERISTIC ON SHOCK LEVEL
RAILROAD FREIGHT TRUCK

NO. OF ACCELEROMETER COUNTS AT STATED INTENSITIES IN 120 MILE RUN

3 - 11/16" TRAVEL SPRINGS - 145,000 LB. CAPACITY
73,000 LB. RAIL LOAD

SHOCK INTENSITY – % G. ACCELERATION	25%	35%	50%	75%
REFERENCE TRUCK – BUILT-IN FRICTION SNUBBER	6187	1689	129	13
CHRYSLER DESIGN TRUCK – SELF-CONTAINED CONSTANT FRICTION SNUBBER	775	219	36	3
PERCENT SHOCK REDUCTION	87.5	87.0	72.0	77.0

3 - 11/16" TRAVEL SPRINGS - 169,000 LB. CAPACITY
145,000 LB. RAIL LOAD

SHOCK INTENSITY – % G. ACCELERATION	25%	35%	50%	75%
REFERENCE TRUCK – BUILT-IN FRICTION SNUBBER	3165	835	8	0
CHRYSLER DESIGN TRUCK – SELF-CONTAINED CONSTANT FRICTION SNUBBER	120	30	0	0
PERCENT SHOCK REDUCTION	96.2	96.4	100	–

SURFACE CURVATURE
A TOOL FOR ENGINEERS

ERNEST WILDHABER*
Rochester, New York

The treatment of surface curvature presented here has been made by an engineer for engineers.

Curvature in general can be based at a chosen point of a given surface. It provides a mathematical means for simplifying a complicated surface in an area around a mean point, and to describe the surface with respect to the tangent plane. The curvature surface is an exact representation of the actual surface only in the infinitesimal area around the chosen point. Beyond this area it is only approximate. However, the information given by values of curvature is relatively simple to obtain, and is sufficient for many practical purposes. For instance, this curvature method can be used in the milling and grinding of helical surfaces existing on gear teeth, worm threads, on the relieved sides of the cutting teeth of many rotary tools, helical gashes of hobs. The profile shape of the grinding wheel is different from the profile of the helical tooth or thread, but its curvature can readily be found by relatively simple equations.

Surface curvature can be used to define warped tooth surfaces and relief surfaces of hobs, as an aid to producing them correctly. It can be used to determine the ease-off at the tooth ends, and to check whether or not a given desired ease-off can be obtained with a given method, and how it can be obtained.

The known mathematical treatment of surface curvature employs plane sections laid in various directions through the surface normal at a given mean point. The treatment presented here proceeds in a more direct manner.

Surface Elements

We shall make continual use of the idea of an elementary cylindrical surface.

Figure 1 shows a cylindrical curvature surface in a view along its straight-line generators. Figure 2 is a corresponding radial view. Its drawing plane is the tangent plane at mean point O of the surface, contacting it along line OX.

A point P, with coordinates x, y in the tangent plane, is at a distance z′ (Figure 1) from the tangent plane. If P is a point of the cylindrical surface with curvature radius R_1, z′ will fulfill the equation

$$z' = \frac{y^2}{2\ R_1} \tag{1}$$

This well-known equation does not need to be proved here, it being understood that all terms in y of higher than the second power do not need to be carried in curvature equations.

The parallel lines shown in Figure 2 are lines of constant elevation from the drawing plane. They constitute a topographical map of the cylindrical surface.

*Copyright by the author.

The inclination i of the surface normal PP′ at P from a line perpendicular to the drawing plane is like the inclination of a radius R_1 to P. It can be obtained by differentiation:

$$\tan i = \frac{d z'}{d y} = \frac{y}{R_1} \qquad (2)$$

At the infinitesimal angles considered, the trigonometric tangent and the arc are interchangeable.

The cylindrical curvature element thus has a simple system of normals. They appear at right angles to line OX in Figure 2. Their inclination i, or tan i, increases in direct proportion to the distance y of a considered point from element line OX. And the elevation z′ from the tangent plane increases with the square of distance y.

A Pair of Component Elements

Figure 3 is a radial view similar to Figure 2, showing the same line of contact OX with the drawing plane, and added to it another line of contact OX′, at an angle Σ to OX. Line OX represents the line of contact of the tangent plane at O with a cylindrical surface of curvature radius R_1. OX′ represents the line of contact of the same plane with a cylindrical surface of curvature radius R_2. A point P of the latter surface has an elevation z″ from the drawing plane. z″ is proportional to the square of distance e of P from OX′, and inversely proportional to the curvature radius R_2.

$$z'' = \frac{e^2}{2 R_2}$$

This is analogous to equation (1).

Let us now consider any curvature surface whatsoever, that touches the drawing plane of Figure 3 at O. Its elevation z from the drawing plane can always be considered the sum of the elevations z′ and z″ of two cylindrical elements.

For proof, we express elevation z from the tangent plane in a quadratic equation:

$$z = C_1 x^2 + C_2 x y + C_3 y^2$$

wherein C_1, C_2, C_3 are numerical coefficients.

There can be no terms simply in x and y, because the elevation z is measured from the tangent plane at the origin of the coordinate system. This equation is the most general form with only quadratic terms.

Distance e, (Figure 3), can be expressed as

$$e = x \sin \Sigma - y \cos \Sigma \ ,$$

$$e^2 = x^2 \sin^2 \Sigma - 2 x y \sin \Sigma \cos \Sigma + y^2 \cos^2 \Sigma$$

Hence $(z' + z'')$ amount to

$$(z' + z'') = x^2 \frac{\sin^2 \Sigma}{2 R_2} - xy \frac{2 \sin \Sigma \cos \Sigma}{2 R_2} + y^2 \left(\frac{\cos^2 \Sigma}{2 R_2} + \frac{1}{2 R_1}\right)$$

This sum $(z' + z'')$ is equal to z when:

$$C_1 = \frac{\sin^2 \Sigma}{2 R_2} \ ; \quad C_2 = - \frac{2 \sin \Sigma \cos \Sigma}{2 R_2} \ ; \quad C_3 = \frac{\cos^2 \Sigma}{2 R_2} + \frac{1}{2 R_1}$$

By division

$$-\frac{2\ C_1}{C_2} = \tan \Sigma \ . \quad \text{Hence}$$

$$\frac{1}{R_2} = \frac{-\ C_2}{\tan \Sigma \ \cos^2 \Sigma} = \frac{C_2^2}{2\ C_1\ \ \cos^2 \Sigma} \ ; \ \frac{\cos^2 \Sigma}{R_2} = \frac{C_2^2}{2\ \ C_1}$$

$$\frac{1}{R_1} = 2\ C_3 - \frac{\cos^2 \Sigma}{R_2}$$

These equations can always be fulfilled. It has thus been demonstrated that the elevations z of any curvature surface from its tangent plane can always be resolved into the sum of the elevations z′ and z″ of a pair of cylindrical elements.

When the general surface and the direction (OX) of one element are given, both elements and their included angle (Σ) can be determined.

By varying the direction at which one element is taken, a given surface can be resolved into an infinite number of pairs of cylindrical elements, of which each pair fully defines the curvature aspects of the surface.

The curvature radii R_1, R_2 of the cylindrical elements may be plus or minus, that is applied from one side or the other to the tangent plane. Figure 4 is a topographical map of a curvature surface where the curvature radii R_1, R_2 of the pair of elements are both positive or botn negative. Here the lines of constant elevation from the tangent plane are ellipses E. Figure 5 is a map of a curvature surface where the two curvature radii R_1, R_2 of the pair of elements are of opposite sign, one being positive and the other negative. Here the lines of constant elevation are hyperbolas H.

Since these curves have an infinitesimal constant elevation from the tangent plane, they correspond to the Dupin indicatrix in surface geometry.

The Surface Normals

It has been shown how a general curvature surface can be resolved into a pair of cylindrical elements, so that its elevations z from the tangent plane are equal to the sum of the elevations of the elements.

And it is also a fact that the inclinations of the normals of the two elements at any point P, being infinitely small angles, can be geometrically added to each other, to give the surface normal as their resultant. This will next be shown.

Let us determine the surface normal at point P, (Figure 3).

OP is at an angle α from OX, and at an angle ($\Sigma - \alpha$) from OX′. It will be referred to as R = OP. Accordingly the distances y of point P from OX and the distance e of P from OX′ amount to

$$y = R \sin \alpha \quad ; \quad e = R \sin (\Sigma - \alpha)$$

By plotting a constant distance on the normals from the points of the curvature surface, these normals can be treated as vectors. We choose a unit distance, that is a distance of one (= 1). It can be plotted at any desired scale. The inclination e of a normal to the direction of the normal at O is then measured by the projection of this distance to the tangent plane. The length of the projection equals sin i. At the infinitesimal angles considered it is also arc i or tan i.

The inclination i, and the projected distance which represents it, can be obtained by geometrical addition of component inclinations of given directions. It can also be resolved into component inclinations of any given directions. We shall refer to these components briefly as components of the normal.

The component PP′ of the normal at P is perpendicular to OX and amounts to

$$PP' = \frac{y}{R_1} = R\frac{\sin\alpha}{R_1}$$

It is the normal of the cylindrical element which extends in the direction OX.

The component PP″ of the normal at P is perpendicular to OX′ and amounts to

$$PP'' = \frac{e}{R_2} = R\frac{\sin(\Sigma - \alpha)}{R_2}$$

It is the normal of the cylindrical element which extends in the direction OX′.

The resultant PQ is obtained by geometrically adding the two components. That is, P′Q is equal in length to PP″ and parallel thereto. The direction of PQ is the direction of the resultant inclination, and the length PQ is a measure of the inclination angle. It measures this angle at the same scale as PP′ measures the component inclination i.

If we resolve the surface into a pair of cylindrical elements, of which one is in the direction of OP, then the other element of the pair is perpendicular to the resultant PQ. For the inclination PQ is produced solely by said other element, since point P lies on element OP.

The resultant inclination PQ of the surface normal at P can also be expressed with respect to the radial plane OP, which contains the surface normal at origin O and the radius vector OP. It can be resolved into a radial component i_r in this plane, and into a peripheral component i_p perpendicular thereto:

$$i_r = PP'\sin\alpha + PP''\sin(\Sigma - \alpha)$$

$$i_r = R\left[\frac{\sin^2\alpha}{R_1} + \frac{\sin^2(\Sigma - \alpha)}{R_2}\right] \qquad (3)$$

$$i_p = PP'\cos\alpha - PP''\cos(\Sigma - \alpha)$$

$$i_p = R.\left[\frac{\sin\alpha\cos\alpha}{R_1} - \frac{\sin(\Sigma - \alpha)\cos(\Sigma - \alpha)}{R_2}\right]$$

$$i_p = (1/2)R.\left[\frac{\sin 2\alpha}{R_1} - \frac{\sin 2(\Sigma - \alpha)}{R_2}\right] \qquad (4)$$

Component i_p becomes zero when

$$\frac{\sin 2\alpha}{R_1} = \frac{\sin 2(\Sigma - \alpha)}{R_2} = \frac{\sin 2\Sigma\cos 2\alpha - \cos 2\Sigma\sin 2\alpha}{R_2}$$

Angle α of this special case will be called α_0. Multiplying both sides with $\frac{R_2}{\sin 2\alpha}$ we obtain:

$$\frac{R_2}{R_1} = \sin 2\Sigma\ \mathrm{ctn}\ 2\alpha_0 - \cos 2\Sigma\ , \qquad \text{and}$$

$$\mathrm{ctn}\ 2\alpha_0 = \frac{\frac{R_2}{R_1} + \cos 2\Sigma}{\sin 2\Sigma} \qquad (5)$$

also $$\mathrm{ctn}\ 2\alpha_0 - \mathrm{ctn}\ 2\Sigma = \frac{R_2}{R_1\sin 2\Sigma} \qquad (5a)$$

Principal Directions

The value of the cotangent is repeated at every half turn. Thus $\operatorname{ctn}(2\alpha_0 + 180^0) = \operatorname{ctn} 2\alpha_0$. Accordingly both angles $2\alpha_0' = (2\alpha_0 + 180^0)$ and $2\alpha_0$ fulfill the above equation. In addition to the solution α_0 there is therefore also a further solution $\alpha_0' = (\alpha_0 + 90^0)$.

In both cases each of the normals lies in a radial plane containing the normal at O, in planes OO′ and OO″ in Figures 4 and 5. The two planes are at right angles to each other. Their directions are known as the principal directions. They are also the directions of a pair of cylindrical elements, into which the surface can be resolved.

Let us now determine the inclination of the normals at points O′ and O″ of the principal direction.

Their peripheral components i_p equal zero, as pointed out above. There is only the radial component i_r at each point. At O′ it is equal to OO′ divided by the curvature radius R_{2_0} of the cylindrical element OO″. With OO′ = OO″ = unity = 1, we obtained with equation (3)

$$i_r' = \frac{1}{R_{2_0}} = \frac{\sin^2 \alpha_0}{R_1} + \frac{\sin^2 (\Sigma - \alpha_0)}{R_2}$$

At O″, i_r is equal to OO″ = 1 divided by the curvature radius R_{1_0} of the cylindrical element OO′. Thus

$$i_r'' = \frac{1}{R_{1_0}} = \frac{\sin^2 (\alpha_0 + 90^0)}{R_1} + \frac{\sin^2 (\Sigma - \alpha_0 - 90^0)}{R_2}$$

$$= \frac{\cos^2 \alpha_0}{R_1} + \frac{\cos^2 (\Sigma - \alpha_0)}{R_2}$$

By addition:

$$i_r' + i_r'' = \frac{1}{R_{1_0}} + \frac{1}{R_{2_0}} = \frac{1}{R_1} + \frac{1}{R_2} \qquad (6)$$

inasmuch as for any angle α, $\sin^2 \alpha + \cos^2 \alpha = 1$.

R_1 and R_2 are seen to be the curvature radii of any pair of cylindrical elements, into which the surface can be resolved. $\frac{1}{R_1}$ and $\frac{1}{R_2}$ are their curvatures. We conclude therefore:

The sum of the curvatures of any pair of cylindrical elements of a curvature surface is constant.

The equations apply also when the radii R_1, R_2 have opposite signs, one being positive and the other negative. In this case R_{1_0}, R_{2_0} also have opposite signs.

Let us consider an axial plane laid through the axis of the cylindrical surface element OO″, Figure 4, and containing the surface normal at say O′. This plane intersects the tangent plane at O in a line parallel to this axis and parallel to OO″. It contains not only the surface normal at O′, but also all other surface normals passing through this line. For their directions can be made up of a component parallel to the surface normal at O′, and a component parallel to the axis and to the line. Both these components lie in the considered axial plane. Thus all normals passing through this line also pass through the axis.

Other axial planes intersect the tangent plane in other lines parallel to the axis. And all the normals passing through these other lines also pass through the axis. This fact applies to each of the two principal directions.

Thus all the surface normals of a curvature surface can be considered to intersect the axes of the two cylindrical surfaces of principal direction.

Line of Contact with a Tangent Cylinder

In the practical application we shall encounter such a problem as the determination of the line of contact between a given curvature surface and a cylindrical surface having the same tangent plane at O, and which has line contact with the curvature surface. In this problem the direction of the cylindrical surface in the tangent plane is given. Another problem is the determination of the curvature radius R_x of this cylindrical surface.

The surface referred to is actually an element of the curvature surface, one of a pair of elements. And the direction of the other element of the pair is the sought direction of the line of contact. The elevations of the curvature surface and of said cylindrical surface from the tangent plane are equal along their line of contact, and they differ increasingly with increasing distance from the line of contact, namely like the elevations of the other element of the pair.

To determine the curvature radius R_x of the cylindrical surface is the same as determining a cylindrical element of given direction. This radius is also the curvature radius of the contour of the curvature surface, when looked at in the given direction along the tangent plane.

A cylindrical surface laid along line OP of the tangent plane is indicated in Figure 4 by its parallel lines of constant elevation from the tangent plane. Our first aim is to determine the line of contact OG. Its construction is best seen in Figure 3.

We know that the sought line of contact will be perpendicular to the direction of the projected normal PQ. For if we consider OP as one element of the curvature surface, this element cannot influence the direction of the normal at P, as it passes through point P. The entire inclination of the normal at P is produced by the other element of the pair, in a direction perpendicular to its lines.

So all we have to do to solve this problem is to determine the direction PQ. The angle $\alpha' = P'PQ$ is the same as the angle $\alpha' = XOG$ between the line of contact OG and OX.

The drawing shows the relationship:

$$\operatorname{ctn} \alpha' = \frac{PP' - P'Q \cos \Sigma}{P'Q \sin \Sigma}$$

The quantities PP' and $P'Q = PP''$ have been determined previously. With these values, the equation becomes

$$\operatorname{ctn} \alpha' = \frac{\frac{\sin \alpha}{R_1} - \frac{\sin (\Sigma - \alpha) \cos \Sigma}{R_2}}{\frac{\sin (\Sigma - \alpha) \sin \Sigma}{R_2}}$$

$$= \frac{\sin \alpha}{\sin (\Sigma - \alpha) \sin \Sigma} \cdot \frac{R_2}{R_1} - \operatorname{ctn} \Sigma \quad ; \quad \text{and}$$

$$(\operatorname{ctn} \alpha' + \operatorname{ctn} \Sigma) = \frac{\sin \alpha}{\sin (\Sigma - \alpha) \sin \Sigma} \cdot \frac{R_2}{R_1} \tag{7}$$

This equation is simplified when referring to the principal directions, which are characterized by $\Sigma = 90^\circ$, and $R_1 = R_{1_0}$, $R_2 = R_{2_0}$:

$$\operatorname{ctn} \alpha' = \tan \alpha \, \frac{R_{2_0}}{R_{1_0}} \quad ; \quad \tan \alpha' = \operatorname{ctn} \alpha \cdot \frac{R_{1_0}}{R_{2_0}} \quad ; \tag{7a}$$

Curvature of Tangent Cylinder

After having determined the inclination α' of the line of contact, we will want to determine the radius of curvature R_x of the tangent cylinder of given direction α.

To this end we consider a point G of the line of contact OG, which is also the direction of the cylindrical element of the pair.

At point G, (see Figure 3), we have the component inclinations GG' and GG". By geometrical addition, these give the resultant GH, made by drawing G'H from G' parallel to GG" and equal thereto.

The determination of the normal at G is quite similar to the determination of the normal at P.

$$GG' = \frac{OG \sin \alpha'}{R_1}$$

GH is perpendicular to OP, as it is solely produced by element OP. It is determined from triangle GG'H with the angles indicated therein. It is

$$GH = \frac{GG' \sin \Sigma}{\sin(\Sigma - \alpha)} = \frac{OG \sin \Sigma \sin \alpha'}{R_1 \sin(\Sigma - \alpha)}$$

GH is also equal to the distance of point G from line OP, divided by the sought curvature radius R_x:

$$GH = \frac{OG \sin(\alpha + \alpha')}{R_x} \quad ; \quad \text{hence}$$

$$\frac{\sin(\alpha + \alpha')}{R_x} = \frac{\sin \Sigma \sin \alpha'}{R_1 \sin(\Sigma - \alpha)}$$

$$R_x = R_1 \frac{\sin(\Sigma - \alpha)}{\sin \Sigma} \cdot \frac{\sin(\alpha + \alpha')}{\sin \alpha'}$$

$$= R_1 \frac{\sin(\Sigma - \alpha)}{\sin \Sigma} \sin \alpha \, [\operatorname{ctn} \alpha' + \operatorname{ctn} \alpha] \; ; \text{ with } \operatorname{ctn} \alpha' \text{ from (7):}$$

$$= R_1 \frac{\sin(\Sigma - \alpha)}{\sin \Sigma} \sin \alpha \left[\frac{\sin \alpha}{\sin(\Sigma - \alpha) \sin \Sigma} \cdot \frac{R_2}{R_1} - \operatorname{ctn} \Sigma + \operatorname{ctn} \alpha\right];$$

$$- \operatorname{ctn} \Sigma + \operatorname{ctn} \alpha = \frac{\sin(\Sigma - \alpha)}{\sin \Sigma \sin \alpha} \quad ;$$

$$R_x = R_2 \frac{\sin^2 \alpha}{\sin^2 \Sigma} + R_1 \frac{\sin^2(\Sigma - \alpha)}{\sin^2 \Sigma}$$

$$R_x \sin^2 \Sigma = R_2 \sin^2 \alpha + R_1 \sin^2(\Sigma - \alpha) \tag{8}$$

and for $\Sigma = 90^0$, $R_1 = R_{10}$, $R_2 = R_{20}$:

$$R_x = R_{20} \sin^2 \alpha + R_{10} \cos^2 \alpha \tag{8a}$$

This is seen to be a very simple result.

The above equation may be further transformed, with

$$\sin^2 \alpha = (1/2)(1 - \cos 2\alpha) \quad ; \quad \cos^2 \alpha = (1/2)(1 + \cos 2\alpha), \text{ to}$$

$$R_x = (1/2)(R_{10} + R_{20}) + (1/2)(R_{10} - R_{20}) \cos 2\alpha, \tag{8b}$$

This equation points to a simple graphical construction. It is shown in Figure 6 for positive values of R_{1_0} and R_{2_0}, where R_{1_0} is larger than R_{2_0}.

$$AB = (1/2)(R_{1_0} + R_{2_0}) \quad ; \quad BC = (1/2)(R_{1_0} - R_{2_0})$$

A circle K is drawn about B with radius BC. It intersects distance AB at C'.

Draw line C'E at an angle α to AB, so that CC'E is angle α. Then angle CBE at the center B of the circle K is equal to 2 α, according to a well-known relationship. Draw EX perpendicular to AC. Distance BX is then BC cos 2 α. It equals the second member of the right side of equation (8b). AB is the first member thereof.

The sum AX equals R_x.

Figure 7 is a similar construction, corresponding to a positive value of R_{1_0} and a negative value of R_{2_0}, numerically smaller than R_{1_0}. Again AX equals R_x. Here the circle K is larger, and reaches beyond A.

It is seen that the curvature radii R_{1_0}, R_{2_0} of the principal directions are a maximum or minimum R .

What occurs when there are more than two cylindrical component elements to define a surface?

More Than Two Elements

If there are three or more cylindrical elements to define a surface, they can always be reduced to a single pair.

Figure 8 is a view of the tangent plane at mean point O. The tangent plane is the drawing plane. Suppose that the three cylindrical elements considered have curvature radii R_1, R_2 and R_3 in their cross-sections and contact the tangent plane in three straight lines all passing through O.

The two elements with curvature radii R_1, R_2 define a surface. This surface, like any other curvature surface, can be resolved into other pairs of elements. Let it be resolved into a pair of elements, of which one element extends in the direction of the element with curvature radius R_3. This direction makes an angle α with OX.

The curvature radius R_x of this one element can be readily determined by equation (8); and the direction α' of the other element of the pair can be computed by equation (7). The curvature radius $R_{x'}$ of the other element can be determined by either equation (8) or equation (6), which can be made to read

$$\frac{1}{R_x} + \frac{1}{R_{x'}} = \frac{1}{R_1} + \frac{1}{R_2}$$

The pair of elements with curvature radii R_x, $R_{x'}$ and angles α and α' define the same component surface as the elements with curvature radii R_1, R_2. To the elevations of this component surface from the tangent plane we now add the elevations of the third element with curvature radius R_3.

As the elements with curvature radii R_x and R_3 have the same direction, they can be combined in a single element.

Let R′ denote its curvature radius. Then its curvature is simply the sum of the curvatures of the two elements:

$$\frac{1}{R'} = \frac{1}{R_x} + \frac{1}{R_3}$$

The surface is then defined by a pair of elements, one having a curvature radius R′ and a direction α, the other having a curvature radius $R_{x'}$ and a direction α'.

Addition of an Element to a Pair of Contacting Surfaces

Let us consider a pair of curvature surfaces which contact along a line passing through mean point O.

The contact conditions remain undisturbed if both of these surfaces are modified equally. A cylindrical element contacting the surfaces at O may be added on the same side of the tangent plane to both surfaces without disturbing the contact. The surfaces altered by the addition of an element continue to contact along a line, a line which extends in the same direction as before.

The curvature in any normal section through O of a resultant surface is the algebraic sum of the curvatures of the original surface and of the added element.

Equidistant or Parallel Surfaces

In gear problems it is often preferable not to work with the tooth surfaces themselves, but with imaginary surfaces equidistant thereto. These may pass through a pitch point in the center of a gear tooth-space. They have the same system of normals as the actual tooth surfaces. And the solution obtained for the equidistant surfaces can be directly translated into the solution for the tooth surfaces.

Before concluding the basic information and taking up applications and examples, a word should be said about conical surfaces and surfaces of revolution.

Conical Surfaces

Curvature does not differentiate between cylindrical surfaces and conical surfaces. Either surface has a straight profile and no curvature in one of the two principal directions; and either is curved in the other principal direction, which we will call the peripheral direction.

On a conical surface the question arises as to the amount of its curvature in the peripheral direction. This curvature is equal to the curvature of a normal section laid in the direction of the periphery. Its radius of curvature differs from the perpendicular distance of the considered point O from the cone axis. It is equal to the normal radius OC in Figure 9:

Figure 9 shows part of a conical surface whose axis O_0C lies in the drawing plane. Figure 10 is a corresponding partial view taken from the left of Figure 9 along the surface normal OC.

In the latter view O and C are projected into the same point. P is considered a point displaced from O in peripheral direction by an infinitesimal amount. Its normal PC continues to pass through C. C is therefore the curvature center of this normal section, and OC is the radius of curvature. The cylindrical element to substitute for the given conical surface has a radius OC equal to the normal radius of the conical surface.

Surfaces of Revolution

Figure 11 shows a surface of revolution whose axis O_0C lies in the drawing plane. The axial section along the drawing plane extends along one of the two principal directions of the surface. The other principal direction is the peripheral direction, as on the conical surface.

The radius of curvature in peripheral direction can be determined exactly as for the conical surface. It is equal to the normal radius OC.

Accordingly the surface of revolution shown can be resolved into two cylindrical surface elements extending in the tangent plane at right angles to each other. One element is convex and the other is concave in the example illustrated. That is, one is positive and the other is negative. The convex element extends along the profile tangent OO_0 and has a curvature radius OC. The concave element extends in a peripheral direction, perpendicular to the drawing plane of Figure 11. Its curvature radius is equal to the curvature radius OC′ of the axial profile.

Summary

We shall now summarize the main ideas of the preceding derivations.

Any curvature surface can be resolved into a pair of cylindrical surfaces or elements, which extend in different directions along the tangent plane at a selected mean point of the surface. They contact the tangent plane along two straight lines which intersect at that point. The curvatures and directions of the two elements are related to one another, as shown in the preceding equations.

The elevations z of any surface point above or below the tangent plane is the algebraic sum of the elevations z′, z″ of the corresponding points of the cylindrical elements.

The surface normals can be treated as vectors, by plotting a constant distance on them. The inclination and direction of a surface normal can be obtained by geometrical addition of the surface normals of the pair of elements, as they appear in a view perpendicular to the tangent plane. This is important, because the system of normals of each cylindrical element is very simple. The normals of an element appear in this view all in the same direction, except when looking directly along the normals, when they are projected into points. And their inclination is directly proportional to the distance of a considered point from the line of contact of the cylindrical element surface with the tangent plane.

The elevation of any point of a cylindrical element from the tangent plane is proportional to the square of said distance.

A curvature surface can be resolved into an infinite number of pairs of component elements, which have different directions, and which include varying angles with each other. Always there is one pair of elements, which are at right angles to each other. Their directions are the principal directions of the curvature surface.

Each element has a curvature radius R and a curvature $\frac{1}{R}$. The sum of the curvatures of the two elements of a pair is constant for all pairs of the same curvature surface.

The curvature of an element is equal to the curvature of the contour of the surface, when viewed in the direction of the element. The line of contact between the curvature and an element extends in the direction of the other element of the pair.

In Figure 4 the element OP contacts the curvature surface along line OG. The surface

itself may be defined by a pair of elements OX and OX′ with curvature radii R_1 and R_2. They include an angle Σ with each other. Equation (7) expresses the relationship between the angles α, α' of OP and OG with base OX.

Equation (8) gives the curvature radius R_x of element OP directly, without reference to the line of contact.

The present treatment of curvature employs tangent cylinders rather than normal sections to define a curvature surface. Problems of tangency can be solved readily therewith.

APPLICATIONS

The Involute Helicoid

In many respects, this is the simplest form of helicoid.

Figure 12 and Figure 13 are a view along the axis of an involute helicoid, and a corresponding front view.

The involute helicoid is made up of straight lines, which are the tangents of its base helix. All points of any of its straight lines have a common tangent plane. Thus the points of straight line OO′ have a tangent plane perpendicular to the drawing plane of Figure 13. The helicoid intersects planes perpendicular to its axis in involutes, of which I, Figure 12, is one.

At any point along a straight-line generator, its curvature surface is the same as on a cone. Thus the points of line OO′ have the same curvature as a conical surface with axis CO′ parallel to the axis of the helicoid. This axis lies on the base cylinder B of the involute helicoid.

OC (Figure 13) is the radius of curvature of the involute helicoid in a section perpendicular to OO′. Dash-and-dot line h is the helix passing through the considered mean point O.

The helix tangent at O makes an angle η with the straight line OO′ of the involute helicoid, which appears enlarged in the axial view, Figure 12.

Let us consider a normal section at O, that is a plane section perpendicular to the helix tangent at that point. It is shown in Figure 14. Point O is in the center of a tooth space. The profile of the considered imaginary helicoid passes through O and appears in dotted lines i. Its curvature center is at C′, on normal OC′. Plane T is parallel to the axis of the helicoid and tangent to the helix at O. The normals (OC′) of opposite sides of the tooth space are inclined at the normal pressure angle φ_n to this plane.

With R denoting the radius of the helicoid to O, the gear radius, and ψ the helix angle at O, distance OC′ can be expressed by

$$OC' = R \frac{\sin \varphi_n}{\cos^2 \psi}$$

and the angle η as

$$\tan \eta = \sin \varphi_n \tan \psi$$

These properties of the involute helicoid are known, and have already been discussed in "Industrial Mathematics."*

*Dr. Werner F. Vogel and Douglas Hughson, "Industrial Mathematics" Volume 3, 1952.

Our immediate object is to determine the profile curvature of a grinding wheel which can produce the involute helicoid. The axis CC_w of the grinding wheel lies in the normal plane, the drawing plane of Figure 14. It has a normal radius OC equal to $\frac{R_c}{\sin \varphi_n}$, wherein R_c is the cutter or wheel radius OC_w. C″ is the center of profile curvature of the grinding wheel, as yet unknown.

Figure 15 shows the tangent plane at O to the imaginary helicoid. The straight line generator OO′ of the involute helicoid is inclined at the above angle η to the direction OX of the helix tangent. The curvature radius R_i (= OC in Figure 13) is known. The quantity $\frac{R_i}{\cos^2 \eta}$ is the radius of curvature OC′ of the normal profile, which extends in the direction OO_0.

The surface of the grinding wheel or cutter can be resolved into a pair of cylindrical elements extending in directions OX and OO_0 respectively. The curvature radius of the latter is equal to OC (Fig. 14) and to

$$\frac{R_c}{\sin \varphi_n}$$

The curvature radius R_p = OC″ of the element extending in direction OX is as yet unknown.

OX and OO_0 are at right angles, so that the use of equation (8a) is in order. Angle η takes the place of angle α. Of the two elements OX and OO_0 one is positive and the other is negative, as their curvatures are opposite. With the present notations, the equation reads

$$R_i = R_p \cos^2 \eta - \frac{R_c}{\sin \varphi_n} \sin^2 \eta \quad ; \quad \text{hence}$$

$$R_p = \frac{R_i}{\cos^2 \eta} + \frac{R_c}{\sin \varphi_n} \tan^2 \eta \quad ; \quad \tan \eta = \sin \varphi_n \tan \chi$$

$$R_p = R \frac{\sin \varphi_n}{\cos^2 \chi} + R_c \sin \varphi_n \tan^2 \chi \qquad (10)$$

The first member of the right side of this equation is the distance OC′ in Figure 14. The second member is distance C′C″, which increases with the size of the grinding wheel.

Thus far we have dealt with the imaginary tooth profile which passes through point O. The actual tooth profile passes through point P. And so does the wheel profile.

C″ remains the curvature center of the wheel profile which passes through P, because this profile lies in a principal direction. Its curvature radius is

$$PC'' = OC'' - OP$$

The wheel profile may be made an involute with curvature center C″ in the case of zero profile ease-off. Its base circle B has its center at C_i and is tangent to OC at C″.

Of course, neither the wheel profile, nor even the normal section of the involute helicoid is an involute exactly. But they are curves of similar character. If OC″ is large, the wheel profile may even be made a circular arc.

Examples (All linear dimensions in inches)

(1) What is the distance OC′ of the center of tooth profile curvature C′ from O of a gear with R = 2 ; φ_n = 20 deg. ; ψ = 45 deg. ?

Answer: $$R \frac{\sin \varphi_n}{\cos^2 \psi} = 1.368$$

(2) How large is distance OC′ when the helix angle is $\psi = 30$ deg. (^{0}) instead of 45 deg. (^{0}) in the above example?

Answer: 0.912

(3) What is the distance of the curvature center C″ (Figure 14) of the wheel profile from the curvature center C′ of the normal profile of the above gear, at a wheel radius $R_c = OC_w = 10$ in.? And what is it at a wheel radius $R_c = 6$ in.?

Answer: It is equal to $R_c \sin \varphi_n \tan^2 \psi$ of equation (10).

$C'C'' = 3.420$ at $\psi = 45^0$; 1.140 at $\psi = 30^0$; $R_c = 10$

$C'C'' = 2.052$ at $\psi = 45^0$; 0.684 at $\psi = 30^0$; $R_c = 6$

(4) What is the distance OC″ of the curvature center C″ from mean point O in the above examples?

Answer: $OC'' = OC' + C'C''$

$OC'' = 4.788$ at $\psi = 45^0$; 2.052 at $\psi = 30^0$; $R_c = 10$

$OC'' = 3.420$ at $\psi = 45^0$; 1.596 at $\psi = 30^0$; $R_c = 6$

(5) What is the curvature radius PC″ of the wheel profile in the above examples, when distance OP is 0.200?

Answer: $PC'' = OC'' - OP$

$= OC''$ (above) $- 0.200$

(6) How much is the curvature radius PC″ of the wheel profile reduced, when the wheel radius R_c is reduced by 1/2 in.?

Answer: $(1/2)\sin \varphi_n \tan^2 \psi$; $\sin \varphi_n = 0.3420$

0.171 at $\psi = 45^0$; 0.057 at $\psi = 30^0$

(7) How large is the curvature radius PC''_b of the wheel profile after this reduction in wheel radius, in the above examples?

Answer: $PC''_b = PC'' - (1/2)\sin \varphi_n \tan^2 \psi$

$PC''_b = 4.417$ at $\psi = 45^0$; 1.795 at $\psi = 30^0$; $R_c = 9.500$

$PC''_b = 3.049$ at $\psi = 45^0$; 1.339 at $\psi = 30^0$; $R_c = 5.500$

(8) How much separation is there between the wheel profiles with curvature radii PC″ and PC''_b, at a point Q corresponding to the top end of the tooth profiles, when PQ = 1/4?

Answer: Because the two wheel profiles are curves similar in character, their separation Δz at point Q is produced principally by the difference in curvature. It is approximately

$$\Delta z = (1/2)\, PQ^2 \left(\frac{1}{PC''_b} - \frac{1}{PC''} \right) = (1/2)\, PQ^2 \frac{PC'' - PC''_b}{PC'' \cdot PC''_b}$$

$\Delta z = 0.00026$ at $\psi = 45^0$; 0.00054 at $\psi = 30^0$; $R_c = 10$ to 9.5

$\Delta z = 0.00054$ at $\psi = 45^0$; 0.00095 at $\psi = 30^0$; $R_c = 6$ to 5.5

A grinding wheel dressed to a constant axial profile would produce a variation of about Δz at the profile ends, as its radius is gradually decreased by 1/2 inch. These amounts are

between 2 1/2 and 9 1/2 tenth of a thousandth of an inch. This is considered too much variation, especially the larger amounts.

This means that the axial wheel profile should not be held constant that long, but should be changed more frequently.

(9) In practice some ease-off will be provided at the profile ends, depending on the manufacturing accuracy, the rigidity of the teeth and their mounting under the given load, the pitch of the gear teeth. Let it be assumed as an example that a profile ease-off between 0.0004 and 0.00025 in. is desired, that a variation of 0.00015 in. is tolerated.

About how much can the wheel radius R_c be decreased from 10 in. and from 6 in. respectively, to obtain a separation of 0.00015 in. of the wheel profiles at point Q in the above examples? PQ = 1/4 in.

Answer: Solve the equation $\left(\frac{1}{PC''_b} - \frac{1}{PC''}\right) = \frac{2\,\Delta z}{PQ^2} = 0.0048$ for PC''_b. Then $(PC'' - PC''_b) = \Delta R_c \sin \varphi_n \tan^2 \psi$; hence

$$\Delta R_c = \frac{(PC'' - PC''_b)}{\sin \varphi_n \tan^2 \psi}$$

The maximum separation $\Delta z = 0.0015$ in. of the wheel profiles and variation of the tooth-profile ends is attained approximately at the following changes ΔR_c of the wheel radii:

$\Delta R_c = 0.289$ at $\psi = 45^\circ$; 0.143 at $\psi = 30^\circ$; $R_c = 10$ at start

$\Delta R_c = 0.143$ at $\psi = 45^\circ$; 0.082 at $\psi = 30^\circ$; $R_c = 6$ at start

The axial wheel profile should then be changed after a decrease of ΔR_c in the wheel radius.

Non-Involute Helicoids

Let us consider a point O of the mean helix, as was done on the involute helicoid. The mean helix, and the normals at all points of the mean helix, are exactly as they are on a coaxial involute helicoid which has the same tangent plane at O. The general helicoid differs from the involute helicoid by rising above or falling below the involute helicoid on both sides of the mean helix.

Figure 15 shows the common tangent plane at O. The involute helicoid contacts this plane along line OO′, which is inclined at the angle η from the direction OX of the helix. The surface normals at points of line OX can be determined solely from the cylindrical element OO′ which represents the involute helicoid. The general helicoid is obtained by adding to this surface element another surface element, that extends in the direction OX of the helix.

The curvature surface at O of the general helicoid can be defined by a pair of elements which make the above angle η with each other. Element OO′ is the element with curvature radius R_i, which represents the involute helicoid. Element OX has a curvature radius R_h in a section perpendicular thereto, which characterizes the departure from the involute helicoid.

General helicoids may also be defined with a given grinding wheel or milling cutter. The problem then is to determine the curvature radius R_h of element OX, and to find the center of curvature of the normal section for the helicoid which passes through O.

This problem is readily solved by comparing the given grinding wheel with a grinding wheel of the same radius R_c, which has a radius of profile curvature OC″ adapted to produce

the involute helicoid. The curvature $\frac{1}{R_h}$ is simply the difference in profile curvature of the last-named wheel and the given wheel. It is $\frac{1}{OC''}$ minus the curvature of the given wheel.

As stated above, we can begin with the involute helicoid, add or subtract a cylindrical element OX to the wheel, to obtain the specified shape on the wheel, and subtract or add the same cylindrical element OX to the helicoid. The resultant surface remains a helicoid because this element is applied in the direction of the helix at O. Element OX has a curvature radius R_h.

The profile curvature in a normal section, $\frac{1}{OC_x}$ in Figure 16, is the profile curvature of the involute helicoid minus the curvature of element OX :

$$\frac{1}{OC_x} = \frac{1}{OC'} - \frac{1}{R_h} \tag{11}$$

This equation also determines $\frac{1}{R_h}$ when OC_x is given.

To determine a grinding wheel or cutter for a given helicoid, we simply subtract $\frac{1}{R_h}$ from the profile curvature $\frac{1}{OC''}$ of the wheel of equal radius OC_w, which can produce a corresponding involute helicoid.

Examples (All linear dimensions in inches)

(1) A helical worm has a radius of R = 1 1/2 to point O, a normal pressure angle $\varphi_n = 20^\circ$ and a helix angle of $\psi = 60^\circ$, that is a lead angle of 30°, at that radius. We consider an imaginary thread surface which passes through point O. If that surface were an involute helicoid, where would the center of curvature C′ of its normal section be? What does distance OC′ in Figure 14 amount to?

Answer: $$OC' = R \frac{\sin \varphi_n}{\cos^2 \psi} = 2.052$$

(2) If the thread surface were an involute helicoid in the above example, at what distance OC″ from O would the center of curvature C″ of the concave wheel profile be, when the wheel radius $R_c = OC_w = 10$ in.?

Answer: $$OC'' = OC' + C'C'' = 2.052 + R_c \sin \varphi_n \tan^2 \psi \text{ , see (10)}$$
$$= 2.052 + 10.261 = 12.313$$

(3) If the thread surface passing through O and defined in (1) is produced with a straight-sided grinding wheel of $R_c = OC_w = 10$, what is the radius of curvature R_h of the element extending along the helix?

Answer: $$R_h = OC'' = 12.313$$

The use of a straight grinding profile is shown in Figure 16, which is a normal section like Figure 14. Here also the drawing plane contains the wheel axis CC_w.

(4) Where is the center of curvature C_x of the normal section of the thread surface produced with a straight-sided wheel and defined in (3)? What is distance OC_x in Figure 16?

Answer: $$\frac{1}{OC_x} = \frac{1}{OC'} - \frac{1}{R_h} = \frac{1}{2.052} - \frac{1}{12.313}$$
$$OC_x = 2.463$$

The above equation may be transformed:

$$OC_x = \frac{R_h \cdot OC'}{R_h - OC'} \; ; \quad R_h - OC' = C'C'' = R_c \sin \varphi_n \tan^2 \psi$$

$$OC_x = OC' \frac{(C'C'' + OC')}{C'C''} = (1 + \frac{OC'}{C'C''}) OC$$

$$OC_x = R \frac{\sin \varphi_n}{\cos^2 \psi} (1 + \frac{R}{R_c \sin^2 \psi}) \qquad (12)$$

Equation (12) permits to obtain OC_x directly.

(5) The actual thread surface is equidistant to the above thread surface and passes through a point P at a distance OP from O. It is further defined by the helix angle $\psi = 60^0$ at O, by the normal pressure angle $\varphi_n = 20^0$ at O, and by the radius R to point O. It is produced with a straight-sided grinding wheel of a radius $R_c = OC_w = 10$ in.

If we were to cut this surface with an end-mill with radial axis OC_w (Figure 16), where would the center of curvature of the cutting profile be?

Answer: At C_x ; because at point O an end-mill corresponds to $R_c = O$.

(6) Does the center of curvature C_x also coincide with the center of curvature of the normal thread-profile which passes through point P?

Answer: Nearly so, but not exactly.

(7) The above thread, which is ground with a straight-sided wheel of 10 in. mean radius, $OC_w = 10$, ordinarily will be cut first. It may be cut with a milling cutter of 1 1/2 in. mean radius, $OC_w = 1\ 1/2$. At what distance $R_{p'}$ from O should the curvature center of its concave cutting profile be, to produce the same thread surface as the grinding wheel?

Answer: An involute helicoid would require a radius of curvature OC″ of the wheel profile passing through O :

$$OC'' = 2.052 + 1.500 \sin \varphi_n \tan^2 \psi = 3.591$$

$$\frac{1}{R_{p'}} = \frac{1}{OC''} - \frac{1}{R_h} \; ; \quad R_{p'} = \frac{R_h (OC'')}{R_h - OC''} = 5.070$$

(8) What is the radius of curvature at P of the actual cutter profile in the above example, when OP equals 0.200?

Answer: $5.070 - 0.200 = 4.870$

Figure 17 shows a thread of straight profile in its normal section. In this case a convex profile with curvature center C_y is required on the grinding wheel. At central point O the relative curvature between the imaginary thread profile and wheel profile is the same as in other forms of threads. Thus OC_y is equal to OC_x of Figure 16, when the data are otherwise the same.

(9) Given a helical surface with straight profile (OP) in a cross-section perpendicular to its axis, see Figure 18. It has a helix angle ψ at O, see Figure 19, which is a view of the tangent plane at O. The helix angle is the complement of the lead angle.

What is the profile curvature of a grinding wheel set to the helix angle at O?

The central plane of the grinding wheel then contains the helix tangent at O.

Procedure: The first step is to define the helical surface by a pair of elementary cylindrical surfaces passing through O. One of these cylindrical surfaces represents the involute helicoid of the same lead and the same tangent plane at O. And the other extends in the direction of the helix tangent at O.

The direction of the first element is the projection OY to the tangent plane of a line equally directed as the axis of the helicoid. This projection is here a line parallel to the axis (C).

The curvature radius of the first element is always the distance from point O to the point of tangency of the surface normal with the base cylinder of the involute helicoid. The radius of the latter is here zero, so that e = OC is the radius of the first element.

The line of contact of the second element with the tangent plane is inclined to OY at the helix angle ψ. Its radius of curvature is such that the resultant line OP is straight.

Point P of the first element has an elevation z′ from the tangent plane

$$z' = -\frac{\frac{1}{2}\,OP^2}{R_i} = -\frac{\frac{1}{2}\,OP^2}{e}$$

The negative sign indicates that this point P is below the tangent plane. But to make OP a straight line it should lie IN the tangent plane. That is z′ has to be cancelled out by the elevation z″ of the other element, so that z′ + z″ = O.

$$z'' = \frac{\frac{1}{2}\,(OP\cos\chi)^2}{R_h} = \frac{\frac{1}{2}\,OP^2}{e} \quad ; \quad \text{Hence}$$

$$R_2 = R_h = e\cos^2\psi$$

$$R_1 = R_i = -e$$

$$\Sigma = (180^\circ - \psi)$$

The grinding wheel surface is defined by the elementary cylindrical surfaces which extend in the two principal directions. Figure 20 shows the grinding wheel in an axial section. R_c is its radius to O. R_p is the radius of curvature of its profile, at O, as yet unknown.

The cylindrical element representing the profile curvature contacts the tangent plane along a line OT (Figure 21) inclined at the helix angle ψ to the vertical OY; along the same line as element R_h of the helicoid. The contact line of the other element of the grinding surface is perpendicular thereto. It is the projection of the wheel axis to the tangent plane. The radius of this other element is R_c.

Line Contact

Of the two contacting surfaces each is defined by a pair of elements. The helical surface is defined by elements R_i and R_h; and the grinding surface is defined by element R_c and by element R_p, whose curvature radius (R_p) is yet to be determined.

Elements R_h and R_p have the same direction, and can be combined. This is done by adding an element $-R_h$ to both surfaces, so that it cancels out element R_h of the helicoid. Of the helicoid only element R_i remains after this operation.

On the other surface the transformed element with direction OT is obtained by adding the curvature of the elements - R_h and R_p, thus

$$\frac{1}{X} = -\frac{1}{e\cos^2\psi} + \frac{1}{R_p}$$

$$X = \frac{-1}{\frac{1}{e\cos^2\psi} - \frac{1}{R_p}}$$

This transformation simply means that the elevations z are now measured from the cylindrical surface R_h, not from the tangent plane.

The known element $R_i = -e$ with direction OY should have line contact with the surface given by its principal elements R_c and X. So we can write equation (8a) as follows:

$$-e = R_c \sin^2\psi + X\cos^2\psi$$

$$= R_c \sin^2\psi - \frac{\cos^2\psi}{\frac{1}{e\cos^2\psi} - \frac{1}{R_p}}, \quad \text{hence}$$

$$\frac{1}{e\cos^2\psi} - \frac{1}{R_p} = \frac{\cos^2\psi}{e + R_c\sin^2\psi}$$

$$\frac{1}{R_p} = \frac{1}{e\cos^2\psi} - \frac{\cos^2\psi}{e + R_c\sin^2\psi} \quad ; \quad \text{by transformation}$$

$$R_p = e\cos^2\psi\,\frac{R_c + e + e\,\mathrm{ctn}^2\psi}{R_c + e + e\cos^2\psi}$$

Numerical example: $e = 1''$, $R_c = 2''$, $\psi = 20^0$

$$R_p = 2.399''$$

(10) What would the profile radius R_p amount to in the above example in the limit case of an infinitely large grinding wheel diameter?

Answer: The grinding surface would then be a cylindrical surface OT.

$$R_p = R_h = e\cos^2\psi = 0.883''$$

(11) What are the principal directions of the helical surface defined in example 9?

Answer: Use equation (5a) to determine $2\alpha_0$. In this equation

$$2\Sigma = 2(180^0 - \psi), \quad -\mathrm{ctn}\,2\Sigma = \mathrm{ctn}\,2\psi, \quad \sin 2\Sigma = -\sin 2\chi.$$

$$R_1 = -e \ , \quad R_2 = e\cos^2\psi \ .$$

$$\mathrm{ctn}\,2\alpha_0 + \mathrm{ctn}\,2\psi = \frac{e\cos^2\psi}{e\sin 2\psi}$$

$$\mathrm{ctn}\,2\alpha_0 = \frac{\cos^2\psi - \cos 2\psi}{\sin 2\psi}$$

$$= \frac{\cos^2\psi - (2\cos^2\psi - 1)}{2\sin\psi\cos\psi}$$

$$\mathrm{ctn}\,2\alpha_0 = \tfrac{1}{2}\tan\psi$$

$$2\alpha_0 = 79^0\,41' \quad ; \quad \alpha_0 = 39^0\,51'$$

The principal directions are indicated in Figure 22, which is a view of the tangent plane. The surface is concavely curved in the direction α_0.

(12) In example (9) the profile curvature of the grinding wheel depends on the wheel diameter. It changes as the wheel diameter is decreased. Is there any way of grinding the helicoid so that the same grinding profile can be used regardless of wheel diameter?

Answer: Yes. When the projected axis of the grinding wheel extends in a principal direction, in direction α_0, the normals at all points of the grinding line lie in the axial plane of the grinding wheel, and intersect the wheel axis. The normals of a given profile remain the same regardless of the distance between the profile and the wheel axis.

The wheel is then set to a helix angle $(90^0 - \alpha_0) = 50^0\,9'$ in the above example.

This procedure can be used when there are no projections at the ends of the straight profile OP (Figure 18).

(13) What is the radius of curvature of the grinding profile in example (12)?

Answer: The radius of curvature R_{p_0} of the grinding profile and wheel profile can be obtained from equation (3), where i_r at α_0 is $\frac{R}{R_{2_0}} = \frac{R}{R_{p_0}}$. Hence

$$\frac{1}{R_{p_0}} = \frac{\sin^2\alpha_0}{(-e)} + \frac{\sin^2(180^0 - \psi - \alpha_0)}{e\cos^2\psi}$$

$$= \frac{1}{e\cos^2\psi}\left[\sin^2(\psi + \alpha_0) - \sin^2\alpha_0\cos^2\psi\right]$$

$$= \frac{1}{e\cos^2\psi}\left[\sin^2\psi\cos^2\alpha_0 + 2\sin\psi\cos\psi\sin\alpha_0\cos\alpha_0\right]$$

$$R_{p_0} = \frac{e\,\mathrm{ctn}\,\psi}{\sin 2\alpha_0 + \tan\psi\cos^2\alpha_0}$$

$$R_{p_0} = 2.293''$$

Large grinding wheels can be used in this case, if desired.

Conclusion

We have described a treatment of surface curvature that is especially suited for engineers. Examples of its application have been presented. The resulting equations are always simple, even though their derivation may be tortuous at times.

There are many more applications.

The author has applied this curvature method in the generating of spiral bevel gears and of hypoid gears, and has used it extensively in developments made at the Gleason Works. The equations now being used for computation are based on it.

This curvature method and its application can shed light into some still-dark corners of the engineering art.

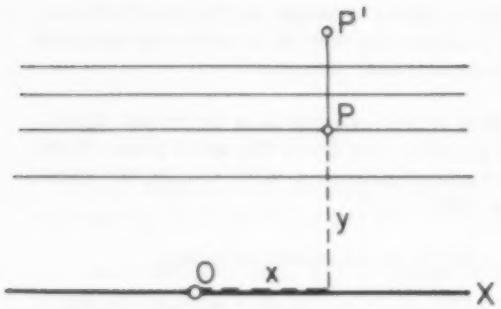


FIG. 2

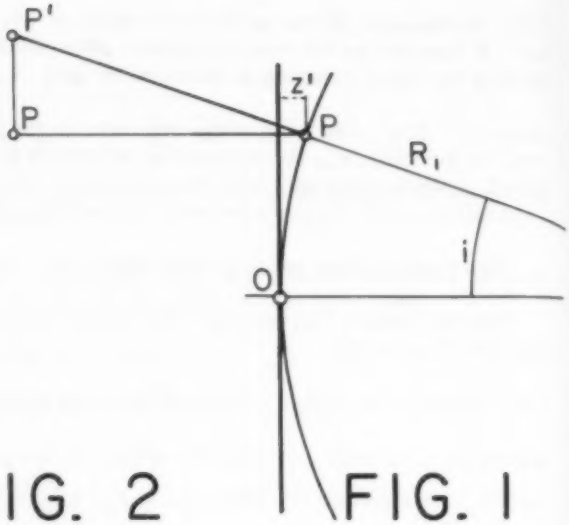


FIG. 1

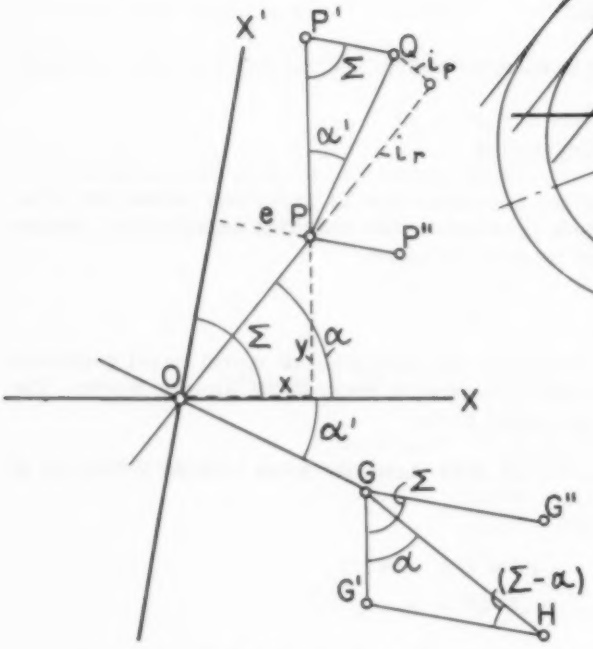


FIG. 3

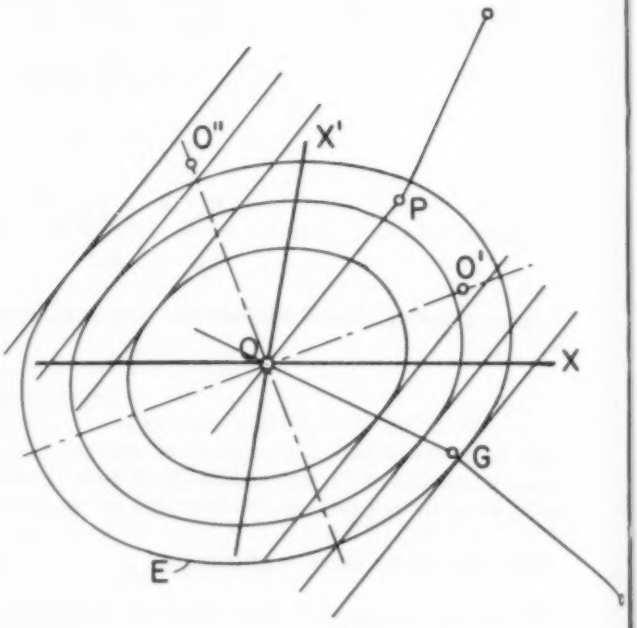


FIG. 4

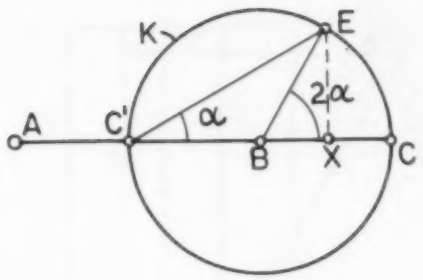


FIG. 6

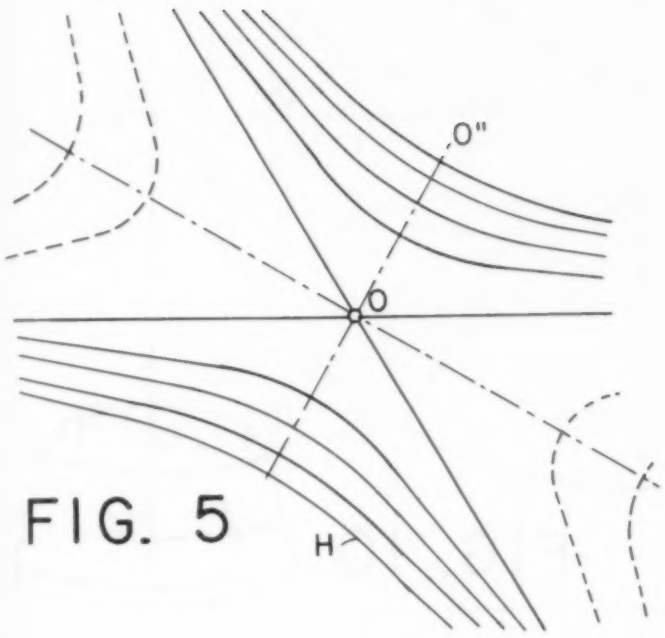


FIG. 5

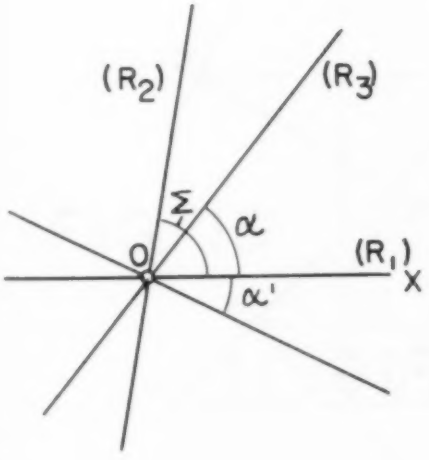


FIG. 8

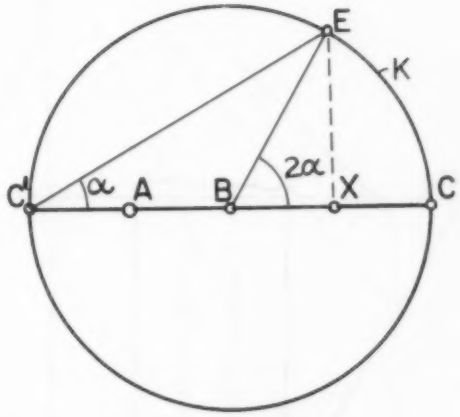


FIG. 7

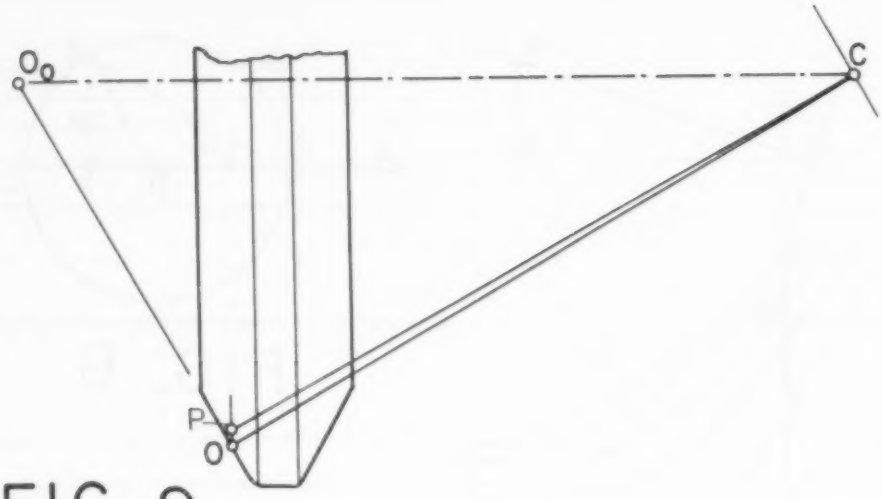


FIG. 9

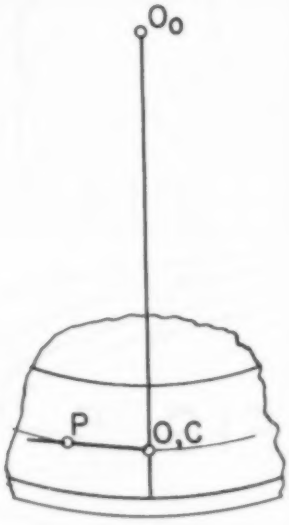


FIG. 10

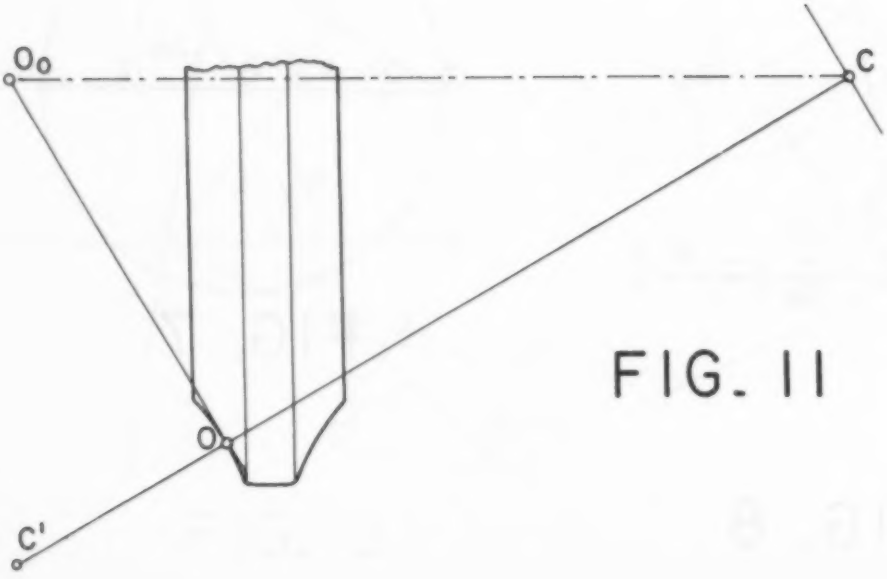


FIG. 11

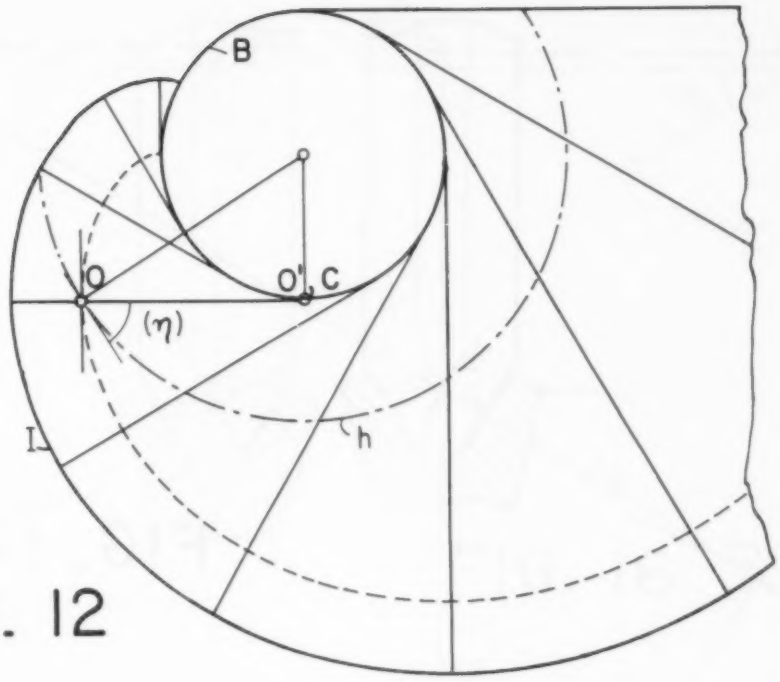


FIG. 12

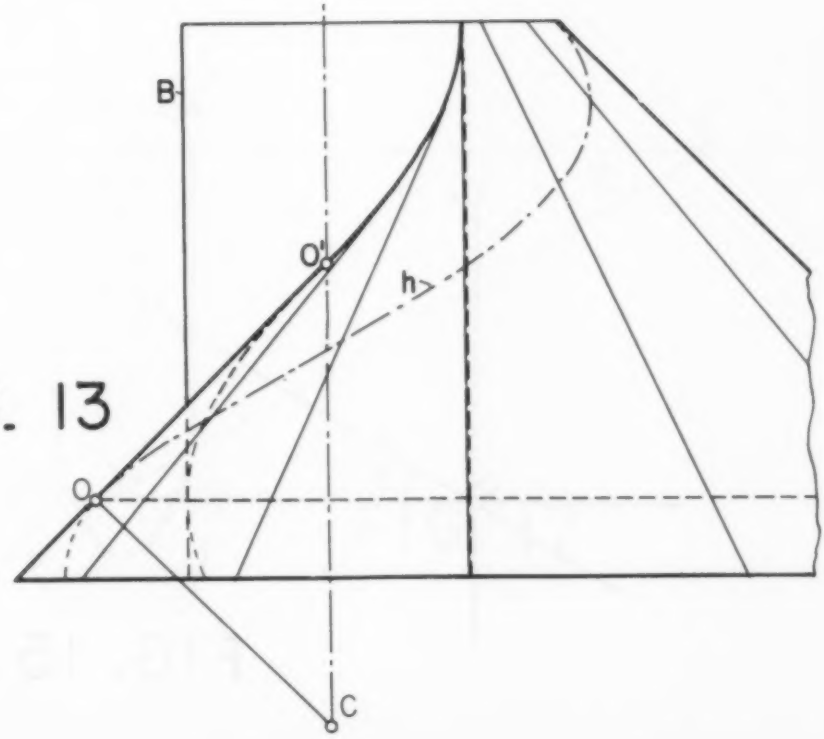


FIG. 13

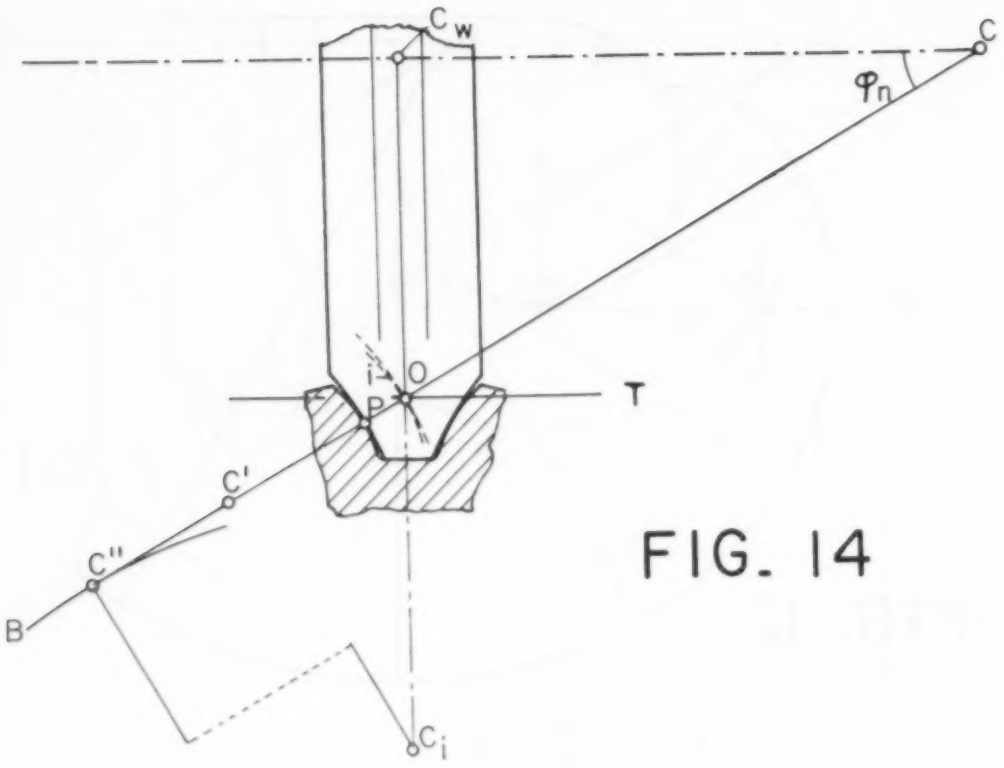


FIG. 14

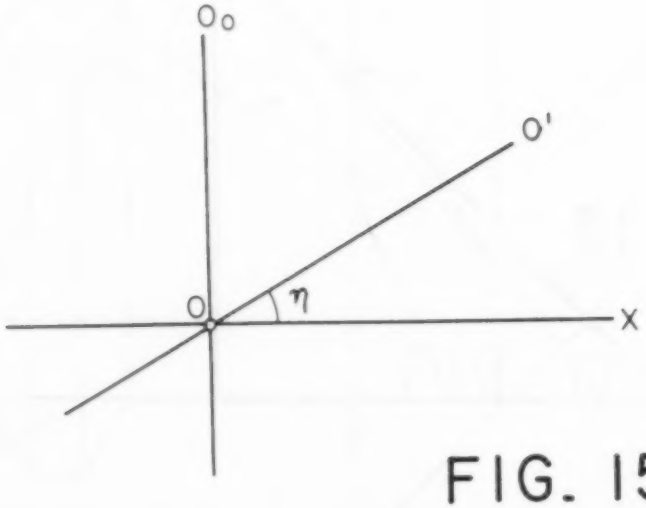


FIG. 15

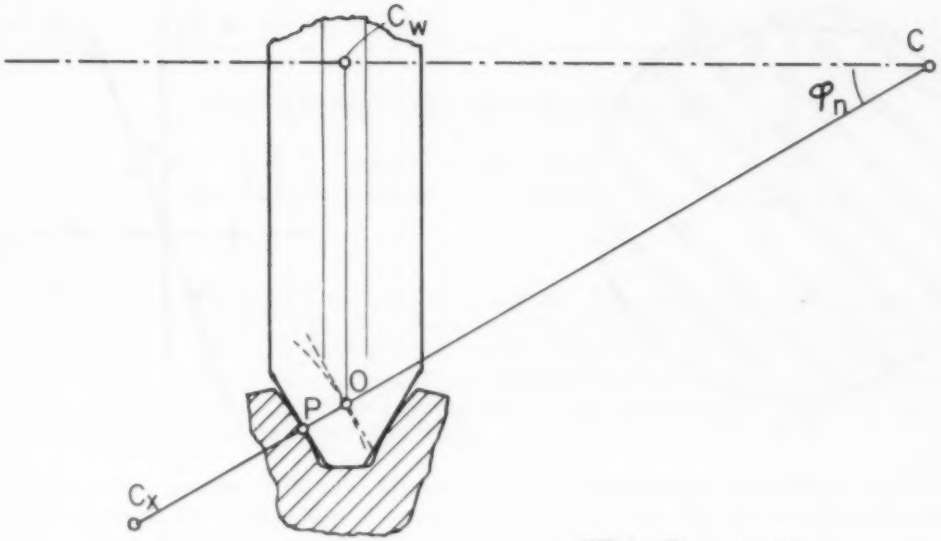


FIG. 16

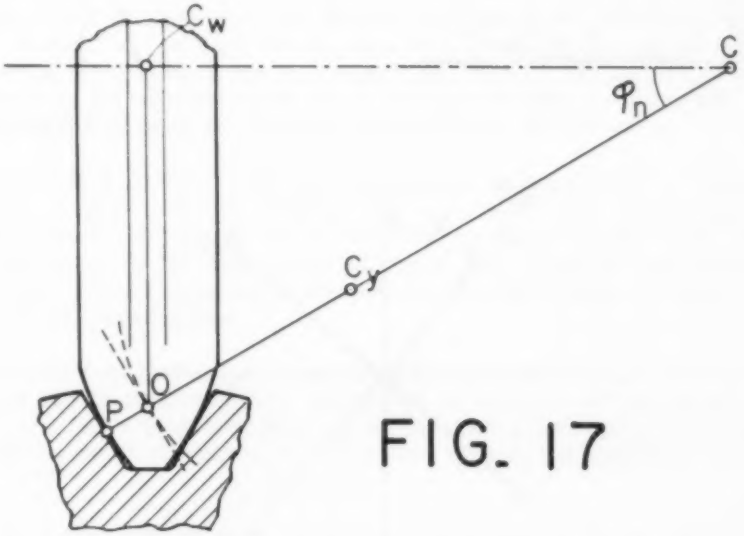


FIG. 17

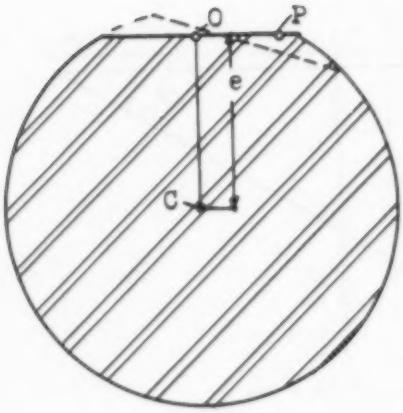


FIG. 18

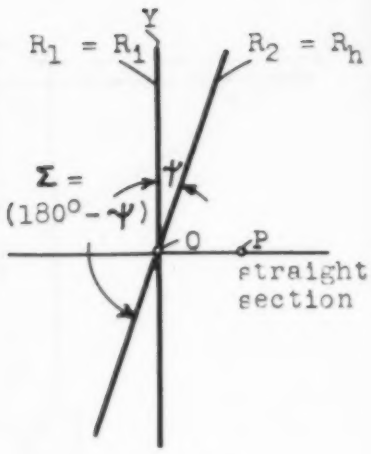


FIG. 19

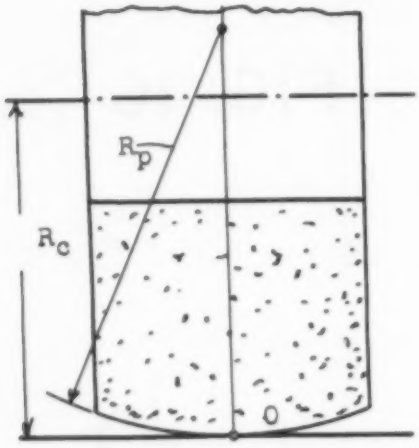


FIG. 20

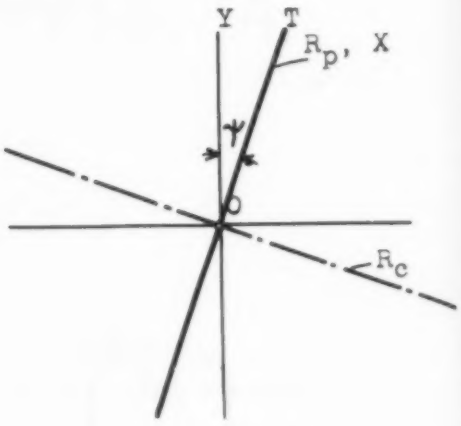


FIG. 21

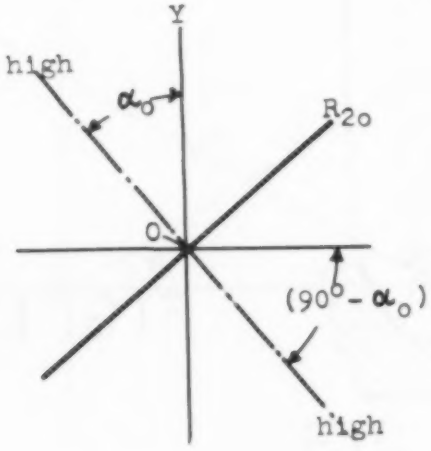


FIG. 22

SOME APPLICATIONS OF SYMBOLIC LOGIC

JAMES C. HETRICK
Continental Oil Company[1]

To those working in industrial mathematics, the last few years have been notable in that many fields formerly thought to be of interest only to pure mathematicians have been brought out of the academies, and have become tools of daily application. The use of matrix methods by economic planners and traffic managers, the application of the theory of linear graphs to the problem of communications within a group; the host of probabilistic models of the operations research groups, all illustrate this trend.

One of the more notable of these newly applied disciplines is that of symbolic logic. Much has been written on the elegant use of the class algebra by Shannon and others in the design of circuits, to the extent that other possible applications of potential importance have been overshadowed. It will not be possible, within the confines of a single paper, to do more than sketch some trivially brief exercises, but it is hoped that these will serve to illustrate the potentialities of the field.

Such a discussion may well begin with the class algebra of Boole, (1) now a century old. The basic principles of this algebra are presented in Table I, using the notation of Birkhoff and MacLane. (2) Here each quantity designates a class, and the quantities defined by operations on classes are themselves classes. The operator " $\cap$," read "and," designates the intersection of classes. Thus $a \cap b$ represents the class all elements of which are elements of both class a and class b. The operator " $\cup$," read "or," designates the union of classes. Thus $a \cup b$ represents the class, any element of which is an element of class a or class b, or both classes a and b. (Note that the Boolean "or" is not the exclusive "or" of ordinary speech.) The symbol " $<$ " is read "is included in" and shows that all elements of the included class are also elements of the including class. Finally the null class of no elements is represented by 0, the universe class which includes all other classes by 1, and the complement of a class, designating all elements not in a class, by a prime.

A convenient method of illustrating the relationships of the algebra of classes is given by use of the Venn diagram. In this diagram, the universe is represented by a square, and classes by areas within the square. The intersection of classes is then represented by overlap of areas and union by the sum of areas. Some Venn diagrams are given in Figure 1, illustrating a class and its complement, the union and intersection of classes, and the distributive laws of the class algebra.

Inspection of the relationships given in Table 1 shows an analogy to ordinary algebra if union is represented by multiplication, intersection by addition, and inclusion by the relation "less than or equal to." This being done, all the ordinary algebraic laws will hold except for formula 10, the second distributive law. For this reason, the algebraic symbols are frequently used.

With this as a preliminary, some illustrative problems involving the use of the class algebra will be given. Statements given without proof may be found in the references cited, and may be illustrated by means of the Venn diagram.

[1]The address on which this paper is based was prepared and delivered while the author was employed by the Ethyl Corporation Research Laboratories.

Illustrative Problem 1

A problem in the construction of hypotheses.

Given three quantities: A, B, and C, which may be independently of any size. It is desired to find the minimum number of hypotheses which are mutually exclusive and totally exhaustive and which will embrace all valid orderings based on equality and inequality.

A table may be constructed showing the equality-inequality relationships by ordered parts. Thus:

A, B	>	>	>	>	>	>	>	>	>
B, C	>	>	>	=	=	=	<	<	<
A, C	>	=	<	>	=	<	>	=	<

etc.

showing 27 possible combinations, of which 14 are invalid or redundant. Thus 13 hypotheses may be used. However, if the quantity N is defined as the class of elements 0, 1, 2, . . . , N in the universe of non-negative real numbers, the possible relationships among the classes A and B are given by

A includes B	$B \cap A' = 0$
B includes A	$A \cap B' = 0$

Therefore only six such relationships are possible, and any hypothesis must include two relations in order to consider all three classes. Thus three hypotheses are sufficient for a totally exhaustive set. The hypotheses are constructed for convenience by pairing on the primed class, and are translated into ordinary algebra. They are then made mutually exclusive by adopting the convention that the equality condition will accompany the first appearance of a pair. The result is:

	Class Algebraic Form	Algebraic Form	Orderings
H.1	$(A' \cap B) \cup (A' \cap C) = 0$	$A \geq B,\ A \geq C$	$A > B > C$ $A > B = C$ $A > C > B$ $A = B > C$ $A = C > B$ $A = B = C$
H.2	$(A \cap B') \cup (B' \cap C) = 0$	$B > A,\ B \geq C$	$B > A > C$ $B > C > A$ $B > A = C$ $B = C > A$
H.3	$(A \cap C') \cup (B \cap C') = 0$	$C > A,\ C > B$	$C > B > A$ $C > B = A$ $C > A > B$

Illustrative Problem 2

A problem in consistency of data (taken from the joint associateship examination, American Actuarial Society, 1935).

Data obtained from a group of 1,000 persons as to race, sex, and marital status were summarized as follows:

number of males	312	number male and married	86
number colored	525	number colored and married	147
number married	470	number male, colored, and married	25
number male and colored	42		

Test this classification for consistency.

TABLE I THE CLASS ALGEBRA

1. $a \cap b = c$
2. $a \cup b = c$
3. $a \cap 0 = 0$
4. $a \cap 1 = 1$
5. $a \cup 0 = a$
6. $a \cup 1 = 1$
7. $a \cup b = b \cup a$
8. $a \cap b = b \cap a$
9. $a \cup (b \cap c) = (a \cup b) \cap (a \cup c)$
10. $a \cap (b \cup c) = (a \cap b) \cup (a \cap c)$
11. $a < b < c \rightarrow a < c$
12. $a \cup a' = 1$
13. $a \cap a' = 0$

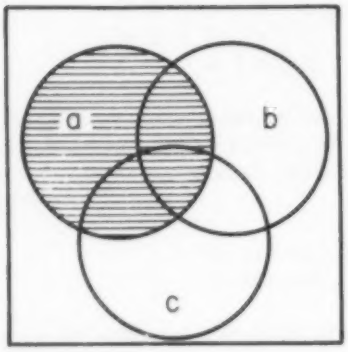
WHERE: a, b, c ARE CLASSES

0	THE NULL CLASS OF NO ELEMENTS
1	THE UNIVERSE CLASS OF ALL ELEMENTS
$\cap$	"AND"
$\cup$	"OR"
$<$	INCLUDED IN
$\rightarrow$	IMPLIES

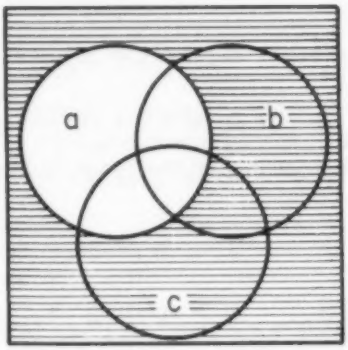
FIGURE I : SOME VENN DIAGRAMS

SHADED AREAS INDICATE THE DEFINED CLASSES

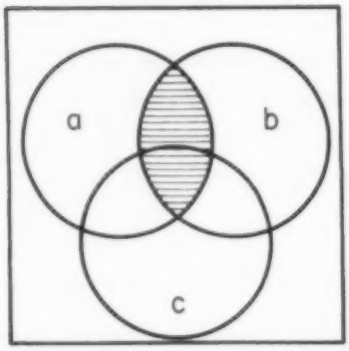
The class a



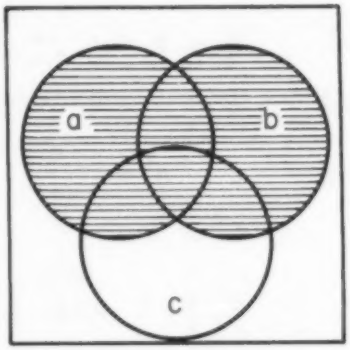
The class a'



The class $a \cap b$

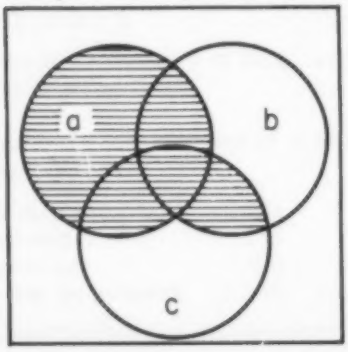


The class $a \cup b$



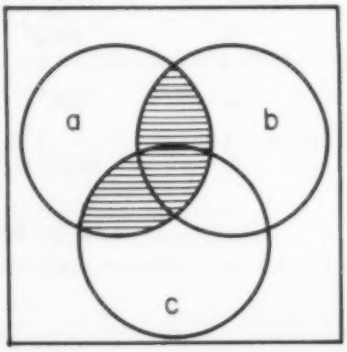
The class

$a \cup (b \cap c) = (a \cup b) \cap (a \cup c)$



The class

$a \cap (b \cup c) = (a \cap b) \cup (a \cap c)$



Class definitions are assigned within the group as: a is the class of male persons, b the class of colored persons, and c the class of married persons. The other classes are then defined as the intersections of these classes. If now we define N(x) to be the number of elements in class x it can be shown, and may be verified on the Venn diagram, that

$$N(a \cup b \cup c) = N(a) + N(b) + N(c) - N(a \cap b) = N(b \cap c) - N(a \cap c) + N(a \cap b \cap c)$$

Placing the proper number values in this expression, we have

$$N(a \cup b \cup c) = 312 + 525 + 470 - 42 - 86 - 147 + 25 = 1057$$

But this is greater than 1,000, the number in the universe of this problem, so that the data are inconsistent.

Illustrative Problem 3

A problem in limitations (taken from Birkhoff and MacLane (2)).

The formula given in the previous problem may often be used, when classification data are incompletely expressed, to establish values, or at least limiting values, for the classes not expressed. For example:

A penny, a nickel, and a quarter are tossed together 100 times. The penny turns up heads on 70 tosses, the nickel heads on 50 tosses, and the quarter heads on 56 tosses. The penny and nickel are heads together exactly 31 times, while the nickel and quarter are heads together 28 times. How many times could all three coins have turned up heads? All three tails?

Using the obvious notation, we have

$$N(p \cup n \cup q) = N(p) + N(n) + N(q) - N(p \cap n) - N(p \cap q) - N(n \cap q) + N(p \cap n \cap q)$$

Substituting the given numbers, and rearranging,

$$N(p \cap n \cap q) = N(p \cup n \cup q) + N(p \cap q) - 117$$

the right hand portion of the equation contains two unknown quantities. However,

$$p \cup q < p \cup n \cup q$$

therefore

$$N(p \cup q) \leq N(p \cup n \cup q)$$
$$N(p \cup q) = \underline{N}(p \cup n \cup q) \quad (\underline{N} = \text{a lower bound})$$

Then

$$\underline{N}(p \cup n \cup q) = N(p \cup q) + N(p \cap q) - 117$$

But it is obvious, and may be verified on the Venn diagram, that

$$N(p \cup q) + n(p \cap q) = N(p) + N(q)$$

Therefore

$$\underline{N}(p \cap n \cap q) = 70 + 56 - 117 = 9$$

So that all three turned up heads at least nine times. Similarly it may be shown that $\overline{N}(p' \cap q' \cap n') = 11$, so that all three coins turned up tails at most eleven times.

Additional illustrative problems could be produced at will to show applications of the class algebra. Enough has been given to show it is capable of detecting contradictions and inconsistencies, manipulating numerical classifications, and studying non-numerical relationships. A fascinating use in the manipulation of complicated verbal conditions, such as those arising in contracts, has been given by Berkeley. (3) I shall return to the class algebra in considering syllogistic reasoning.

First let us consider a very similar algebra, the algebra of propositions, the rules of which are given in Table 2. The similarity to Table 1 is obvious. In fact, if disjunction is written for union, and conjunction for intersection, the rules of the one apply to the other.

Before considering an application of the algebra of propositions, we wish to study a powerful tool known as the truth table. In the construction of a truth table, the problem is to find whether or not a statement is true for all combinations of the truth values of the components. If the sentence has n components, each two-valued, it is itself two-valued and must be evaluated for 2^n combinations. The simplest possible truth table is

p	p′
T	F
F	T

read "If p is true, then p′ is false; and if p is false, then p′ is true." This is self-evident, since the negation of a true statement is false, and the negation of a false statement is true. The truth table for the conjunction $p \cdot q$, read "both p and q," is constructed with the statement being true only if both component propositions are true; whereas the disjunction $p \vee q$, read, "p or q or both," is false only if both propositions are false. Thus:

p	q	p·q	p∨q
T	T	T	T
T	F	F	T
F	T	F	T
F	F	F	F

With this table it is easy to construct more complex tables.

To illustrate the use of a truth table, consider a simple statement, such as "either a man is alive or he is not." It is easily seen that the statement, while not especially informative, is quite true. Any statement of the form "either p or not p," where p is some meaningful proposition, is equally true. But the truth in such cases is dependent not on content but on structure, and such a statement is known as a tautology. It is obvious that the distinction is important and basic.

Consider now the more complex statement: "Either a man is a mathematician and does not drink whiskey in the morning; or he is not a mathematician and doesn't like Mozart; or else he either drinks whiskey in the morning or likes Mozart." It may be of importance to know whether such a statement has significant meaning or is merely an elaborate tautology.

To investigate this, let

p ≡ a man is a mathematician
q ≡ a man drinks whiskey in the morning
r ≡ a man likes Mozart.

TABLE 2 THE ALGEBRA OF PROPOSITIONS

1. $p \vee q = r$
2. $p \cdot q = r$
3. $p \vee 0 = p$
4. $p \cdot 1 = p$
5. $p \vee q = q \vee p$
6. $p \cdot q = q \cdot p$
7. $p \cdot (q \vee r) = (p \cdot q) \vee (p \cdot r)$
8. $p \vee (q \cdot r) = (p \vee q) \cdot (p \vee r)$
9. $p \vee p' = 1$
10. $p \cdot p' = 0$
11. $p \neq q$
12. $(p = 1) \vee (p = 0)$

WHERE: p, q, r ARE CATEGORICAL STATEMENTS

$\vee$ DISJUNCTION $p \vee q \equiv$ p, OR q, OR BOTH

$\cdot$ CONJUNCTION $p \cdot q \equiv$ BOTH p AND q

1 INDICATES TRUTH

0 INDICATES FALSIFY

FIGURE 2 : THE CATEGORICAL STATEMENTS

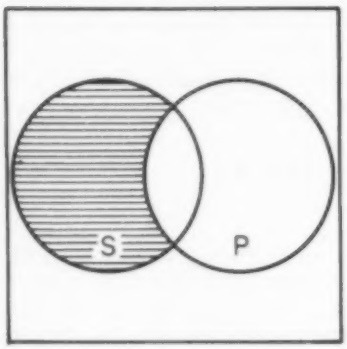
≡ THE CLASS IS EMPTY

|||| THE CLASS IS NOT EMPTY

A
Universal Affirmative

All S is P

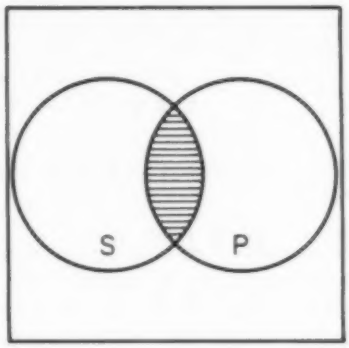
$S \cap P' = 0$



E
Universal Negative

No S is P

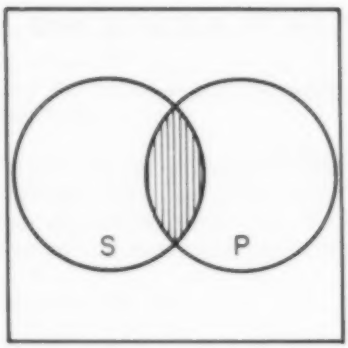
$S \cap P = 0$



I
Particular Affirmative

Some S is P

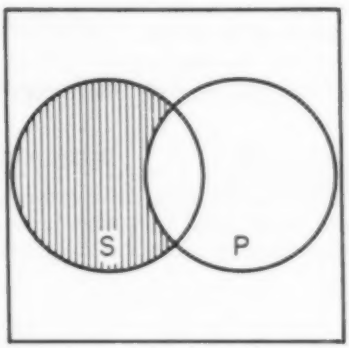
$S \cap P \neq 0$



O
Particular Negative

Some S is-not P

$S \cap P' \neq 0$



Then the complex statement previously given can be represented as:

$$(p \cdot q') \vee (p' \cdot r') \vee (q \quad r)$$

and a truth table built up as follows:

p	q	r	p′·r′	p′·r′	(p·q′) ∨ (p′·r′)	q ∨ r	((p·q′) ∨ (p′·r′) ∨ (q ∨ r)
T	T	T	F	F	F	T	T
T	T	F	F	F	F	T	T
T	F	T	T	F	T	T	T
T	F	F	T	F	T	F	T
F	T	T	F	F	F	T	T
F	T	F	F	T	T	T	T
F	F	T	F	F	F	T	T
F	F	F	F	T	T	F	T

The final column shows that the statement is always true, and is a tautology, giving us no knowledge of the habits of mathematicians.

A truth table can always be built up for any statement, no matter how complex. The table for a polynomial constructed by disjunction, as above, tells us whether a statement has content or only form. A similar polynomial constructed by conjunction is equally important. If the truth table for the conjunction contains a single "T" then the conjoined statements form a consistent group. And in general, the conditional truth of propositions may be so demonstrated.

Application to the Syllogism

Potentially one of the most important applications of the algebra of classes that of manipulating syllogistic reasoning. It will be seen that all of the Aristotelian logic of the categorical syllogism, which in its traditional form has been the stumbling block of students through the ages is reducible to very simple manipulative procedures.

Figure 2 shows the traditional statement of the four types of categorical statements, together with the Boolean equivalents and an illustrative Venn diagram. In a valid syllogism, the two premises are statements in these four forms, and the Boolean product is the conclusion. Valid class algebraic operations on these forms are:

a. Commutativity of classes
b. Completion (if the subject class is not empty)
   thus $S = S \cap (P \cup P')$
   $= (S \cap P) \cup (S \cap P') \neq O$
   $(S \cap P') = O \rightarrow (S \cap P) \neq O$
c. Commutativity of propositions.

By means of these, all forms of immediate inference may be accomplished, and from this all forms of reduction to the first figure follow. This is demonstrated in Table 3. In addition, the universal negative may be rephrased as a universal affirmative and have the same form. Thus:

E	No S is P	$S \cap P = O$
A	All S is P′	$S \cap (P') = S \cap P = O$

Similarly for the particular negative

O	Some S is-not P	$S \cap P' \neq O$
I	Some S is P′	$S \cap P' \neq O$

TABLE 3 CLASS ALGEBRA FORMS OF IMMEDIATE INFERENCE

	A	E	I	O
ORIGINAL	$S \cap P' = 0$	$S \cap P = 0$	$S \cap P \neq 0$	$S \cap P' \neq 0$
CONVERSE	$P \cap S \neq 0$	$P \cap S = 0$	$P \cap S \neq 0$	———
OBVERSE	$S \cap P' = 0$	$S \cap P = 0$	$S \cap P \neq 0$	$S \cap P' \neq 0$
PARTIAL CONTRAPOSITIVE	$P' \cap S = 0$	$P' \cap S \neq 0$	———	$P' \cap S \neq 0$
FULL CONTRAPOSITIVE	$P' \cap S = 0$	$P' \cap S \neq 0$	———	$P' \cap S \neq 0$

From this it follows that

(1) All valid forms may be reduced to the first figure.

(2) In the first figure, only the affirmative forms need be considered.

(3) AAI and AAO in the first figure are equivalent (O → I), other weak cases are reduced to them. They in turn may be derived from AAA, for if $A \cap C' = O$ and $A \neq O$, then $A \cap C \neq O$ as previously shown.

Then the whole of Aristotelian logic may be expressed in the forms:

Theorem I

If $a \cap b' = O$
and $b \cap c' = O$
then $a \cap c' = O$ $(a \cap c \neq O,\ a \neq O)$

Theorem II

If $b \cap c' = O$
and $a \cap b \neq O$
then $a \cap c \neq O$

recognizable as the classical forms Barbara and Darii.

On this subject, Mrs. Langer has written: (4) "It may be seen that the categorical syllogism is a small part of the algebra of classes, namely a sub-system of three terms and their respective complements, and setting forth the relation of inclusion, partial or total, among them. Thousands of men, through thousands of years, have had millions of headaches over

the 24 valid and 40 invalid combinations of these terms: arranging, relating, and naming them. Symbolic logic proves them all equivalent to just three forms of a much greater system."

The utility of these theorems may be shown by considering a simple example, due to Dodgson. (5) We are required to find the conclusions, given the four statements:

(1) No boys under 12 are admitted as boarders.
(2) All the industrious boys have red hair.
(3) None of the day boys learn Greek.
(4) None but those under 12 are idle.

Here we define the universe as being boys admitted to this school, with classes being

a ≡ boys 12 and over
b ≡ boys admitted to this school as boarders
c ≡ industrious boys
d ≡ boys having red hair
e ≡ boys learning Greek

Then the given statements are expressed:

$$a' \cap b = O \qquad (1)$$
$$c \cap d' = O \qquad (2)$$
$$b' \cap e = O \qquad (3)$$
$$a \cap c' = O \qquad (4)$$

Since all are of the form A (universal affirmative) Theorem I may be applied successively. Then

from (1) and (4)	$b \cap c' = O$	(5)
from (5) and (2)	$b \cap d' = O$	(6)
from (6) and (3)	$d' \cap e = O$	(7)

The net conclusion then, from (7) is "All boys learning Greek at this school have red hair." Intermediate conclusions are given by (5) and (6).

More elaborate and non-trivial examples are available in the references cited. (1), (5). The utility of the method is obvious — the task of drawing all valid conclusions from a given set of premises can be rather formidable when using tranditional methods.

The classical logic was greatly expanded by adding the hypothetical and disjunctive syllogisms to the categorical syllogism. Although to show it would expand this paper inordinately, it may be remarked that, just as the class algebra reduces the categorical syllogism to a simple set of relationships, so the algebra of propositions, and the truth table, greatly simplifies the elaborate logical apparatus of the hypothetical and disjunctive syllogism.

We might turn next to the consideration of how the mechanics of handling this powerful tool might be simplified. I wish to introduce a modified notation in which, as before, letters represent classes, but now the complement of a class will be represented by the negative of the class designation. Thus the complement of a is -a. Inclusion will be represented by the algebraic sign of inequality. Now let the universal categorical statements be represented by:

A	All S is P	$S < P$
E	No S is P	$S < -P$

I further propose that these relations be treated as ordinary algebraic inequalities. Thus, from

$$S < P$$

it follows

$$-S > -P$$
$$O > S - P$$
$$O < P - S$$

The problem of reducing sorites of universal propositions then becomes only a matter of algebraic addition. The previous example becomes

$$a < -b$$
$$c < d$$
$$-b < -e$$
$$-a < c$$
$$\overline{O < d - e \rightarrow e < d}$$

which is the conclusion previously reached.

A fairly elaborate set of statements may be rather quickly reduced by this technique. Thus, consider the following set of ten premises, also from Dodgson (loc. cit.):

1. The only animals in this house are cats.
2. Every animal is suitable for a pet, that loves to gaze at the moon.
3. When I detest an animal, I avoid it.
4. No animals are carnivorous, unless they prowl at night.
5. No animals ever take to me except what are in this house.
7. Kangaroos are not suitable for pets.
8. None but carnivora kill mice.
9. I detest animals that do not take to me.
10. Animals that prowl at night always love to gaze at the moon.

We define the universe as being animals, with classes being as follows:

a ≡ animals which are in this house
b ≡ animals which are cats
c ≡ animals which are suitable for pets
d ≡ animals which love to gaze at the moon
e ≡ animals which are carnivora
f ≡ animals which prowl at night
g ≡ animals which take to me
h ≡ animals which are kangaroos
i ≡ animals which kill mice
j ≡ animals which I detest
k ≡ animals which I avoid

The statements are then expressed as:

1. $a < b$
2. $d < c$
3. $j < k$
4. $e < f$
5. $b < i$
6. $g < a$
7. $h < -c$
8. $i < e$
9. $-g < j$
10. $f < d$

$\Sigma = \overline{h < k} \equiv$ the conclusion

from which the conclusion is: "I always avoid kangaroos."

Actually one should not speak of the conclusion in a sorites. There may be many intermediate conclusions, drawn from some of the premises, which are equally valid with the final conclusion, drawn from all the premises. This suggests that instead of a simple summation, the inequalities be ordered. The above becomes:

$$-k \underset{3}{<} -j \underset{9}{<} g \underset{6}{<} a \underset{1}{<} b \underset{5}{<} i \underset{8}{<} e \underset{4}{<} f \underset{10}{<} d \underset{2}{<} c \underset{7}{<} -h$$

where the numbers show the premise from which the inequality was derived. Then all valid statements about the inclusion relationships among the classes may be written systematically by considering each class in turn as being included in each of the classes lying above it in the inequality. The above inequality, then, represents the ten premises and forty-five valid conclusions.

This scheme may be extended to include the particular statements by adopting a symbol of partial inclusion. I shall represent this by a starred inequality sign. The conventions for use of the chair of inequalities now become:

1. The inequality sign shows a relationship which can be operated on with normal algebraic manipulations. The starred inequality represents a relationship which may not be operated on except by transition.

2. Two classes lying in the inequality cannot be related if more than one starred inequality lies between them; they can be related by a starred inequality if only one such starred inequality lies between them; they can be related by a full inequality if only full inequalities lie between them.

Thus in the above example the first statement might be made particular rather than universal, reading, "Some of the animals in this house are cats," and the fourth statement might similarly be weakened to read, "Some carnivora prowl at night." If either of these is done, 45 conclusions may be drawn, but in either case 27 of them will be particular, having come from weak cases of the syllogism. If both changes are introduced, the number of valid conclusions drops to thirty, only eight of which are in universal form. It is obvious that the introduction of particular categorical statements considerably weakens the structure of a sorites. (The above rules are equivalent to the rules for the construction of a valid syllogism in Aristotelian logic as given in, for example, Cooley. (6))

Reflection on this scheme will show that it can be used to group categorical statements into compatible, completely ordered sets, to detect inconsistencies, etc. Since the process involved is essentially that of ordering, it is not too optimistic to hope that a coding system can be devised which will enable the digital computer to sort tremendous masses of categorical statements, and to detect inconsistencies, point out valid conclusions, and otherwise render the services of a logician.

Actually, tremendous advances in this direction have already been made. Much can be done by the calculation of truth tables, and one machine, the Kalin-Burkhart Logical-Truth Calculator, was early built for this particular purpose. (7) More recently the Burroughs Truth-Function Evaluator has been constructed and is able to evaluate truth tables for 10 variables. (8) A general purpose digital computer, the California Digital Computer, has also been coded for truth table computation in a manner that seems adaptible to other computers. (9) And there exists a partially coded program for the ElectroData computer which is capable of finding and printing all valid conclusions for syllogisms with as many as 26 terms and 75 premises. (10)

It has been possible, in this brief paper, to mention a very few of the possible applications of a somewhat neglected technique. It is my hope that with the development of computer techniques further work in applications will be stimulated until the class and propositional algebras will be as commonly utilized a tool as the algebra of quantities.

REFERENCES

1. Boole, George, "An Investigation of the Laws of Thought," 1854. Reprinted by Dover Publications, Inc., New York, 1953.
2. Birkhoff, G., and MacLane, S., "A Survey of Modern Algebra," rev. ed. Macmillan, 1953.
3. Berkeley, E. C., "A Summary of Symbolic Logic and Its Practical Applications," Edmund C. Berkeley and Associates, New York, 1951.
4. Langer, S. K., "An Introduction to Symbolic Logic," 2nd ed., Dover Publications, Inc., 1953.
5. Dodgson, Charles, "Symbolic Logic"; quoted in The Complete Works of Lewis Carroll, Modern Library.
6. Cooley, J. C., "A Primer of Formal Logic," Macmillan, 1942.
7. Berkeley, Edmund C., "Giant Brains," 1949.
8. Burks, A. W., Warren, D. W., and Wright, J. B., MATC VIII, 53 (April, 1954).
9. Abbott, W. R., MATC V, 120 (July, 1951).
10. Reiss, R. F., Private Communication.